Haaga-Helia
University of Applied Sciences

Developing and implementing a usability evaluation framework for Lioncrypt 1.0 software

Elizaveta Kantina

Bachelor's Thesis
Degree Programme in
Business Information
Technology
2016

Author
Elizaveta Kantina

Degree programme
Business Information Technology

Thesis title	**Number of pages and appendix pages**
Developing and implementing a usability evaluation framework for Lioncrypt 1.0 software	**66 + 16**

This thesis presents a framework for usability evaluation of Lioncrypt 1.0 software, developed by Ferus Bestia Oy. Two main objectives are formulated for this thesis: the first objective is to design a framework and the second objective is to use the framework to assess the usability of Lioncrypt 1.0.

Firstly, literature research enabled to study the main usability concepts and investigate the existing trends in usability evaluation. Traditional in-lab usability testing has been found the most suitable in terms of cost and resources. This approach was complemented by the combination of methods and techniques, such as concurrent thinking aloud (CTA), post-task Single Ease Question (SEQ) and post-test System Usability Scale (SUS) questionnaire.

Secondly, scenario-based usability testing was performed during the period from 25.03.2015 to 02.04.2015 in a conference room of Ferus Bestia Oy set up as an informal lab. Usability testing process consisted of 5 independent test sessions, individual for each participant. Screen capture, logging, note taking and audio recording allowed to preserve the raw data. Both quantitative and qualitative data were collected in order to provide a deeper insight and the better coverage of usability issues.

The investigation exposed 12 usability problems across two user interfaces of Lioncrypt 1.0. Potential solutions to fix the identified issues were also provided. The study findings revealed that the usability of Lioncrypt 1.0 is slightly below the industry average level. Among positive findings, it is worth mentioning that learnability achieved a good level. Also, the majority of the participants managed to complete the most of the scenarios without assistance.

The research findings demonstrate that the developed framework successfully serve the purpose of usability evaluation. Therefore, it should be incorporated into the iterative design process. Moreover, it can be adapted to evaluate the usability of software, regardless of technology.

Keywords
Usability, Usability evaluation, Usability testing, Think aloud protocol, SUS, SEQ

Table of contents

List of Figures

Tables

List of Abbreviations

CEO	Chief Executive Officer
CFO	Chief Financial Officer
CTA	Concurrent Think Aloud
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
MVP	Minimum Viable Product
NIST	National Institute of Standards and Technology
OBS	Open Broadcaster Software
PC	Personal Computer
RTA	Retrospective Think Aloud
SEQ	Single Ease Question
SUS	System Usability Scale
UI	User Interface

1 Introduction

The concept of usability has evolved from labeled as “questionable” into an indispensable component of software development cycle. The quality of interaction between user and interface defines the software success on the market. Increasing the level of usability can have a positive influence on the attitude of the customers towards the artifact.

Nevertheless, there is still no total agreement on what usability comprised and which attributes are related to this multifaceted concept. Also, a relationship between usability and user experience remains subject of debate.

Usability is often associated with ease of use (Nielsen 2012d). However, the formal ISO (9241-11 1998; 25063 2014) definition gives the more precise explanation of usability: “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

As can be seen from the above definition, usability encompasses three main dimensions: effectiveness (relates to accuracy), efficiency (refers to speed) and satisfaction (refers to feelings and opinion) (9241-11 1998; 25063 2014). Among other proposed dimensions, special attention should be paid to learnability, as the most mentioned addition to the core attributes of usability (Nielsen 2012d; Quesenbery 2004, 5). The current research utilizes the definition of learnability offered by Nielsen (2012): “How easy is it for users to accomplish basic tasks the first time they encounter the design?”

For the most part, usability is considered as an essential facet of user experience. Speaking of user experience, it appeared to be even more vaguely defined. The literature review provides evidence that user experience more or less embodies every aspect of the interaction between the user and the specific user interface or product. This assumption makes user experience the topic which is outside the scope of this thesis.

With regard to usability evaluation, a significant number of methods have been developed to identify usability issues and assess usability. The most common and reliable of them will be discussed thoroughly in this thesis.

During the my internship in Ferus Bestia Oy, the fact that usability evaluation of the Lioncrypt 1.0 (which is described in more detail in chapter 2) has not been performed, attracted my attention. Even though the development team of Ferus Bestia Oy utilized a

number of modern technologies while developing the Lioncrypt 1.0 software, the usability evaluation seemed to be disregarded.

As a matter of fact, the importance of usability evaluation was admitted by the development team of Ferus Bestia Oy. However the actual research was postponed due to the lack of time, knowledge and sources. The present study intends to fill this research gap.

1.1 Aim, objectives and research questions

The *aim* of this investigation is to develop a framework for usability evaluation of Lioncrypt 1.0. To achieve this aim, two main *objectives* has been set:

- Based on literature research, construct the framework suitable for evaluation of Lioncrypt 1.0.
- Perform usability evaluation study of Lioncrypt 1.0 based on the designed framework.

In pursuance of the objectives above, the following *research questions* were formulated:

- What kind of usability framework would be suitable for evaluation of Lioncrypt 1.0 software?
- What is the degree of usability of Lioncrypt 1.0 software?
- Is the proposed framework applicable for evaluation of Lioncrypt 1.0 software?

1.2 Structure of the thesis

The thesis is divided into seven chapters. Chapter 1 presents the introduction and motivation to the study. Next section introduces thesis aims, objectives, and associated research questions. Chapter 2 gives a brief introduction of the case company and the product.

Chapter 3 and 4 provide the theoretical foundation for the research. In chapter 3 the core theoretical concepts and the background of the research is established, the concepts related to this research are discussed in more detail. In chapter 4 types and the most common methods for usability testing are described and compared.

Chapter 5 begins with stating the objectives for the empirical part of the thesis and providing the usability evaluation framework rationale. It describes the procedure, participants, technical and physical environment. In chapter 6 the results of the empirical study are presented in order following the objectives formulated for the usability evaluation. The chapter closes with a discussion of validity and reliability of the results.

Chapter 7 highlights the significance of the study, summarizes the research and further elaborates the implications of the findings from the previous chapter. Additionally, the chapter provides future research suggestions and recommendations.

2 Organization and product for the report

This chapter provides the brief overview of the company and the product under consideration. The first section presents the general information about the commissioning party. The second section contains product details and description of the intended users.

2.1 Organization info

Finnish start-up company Ferus Bestia Oy was founded in September 2012. Located in Lauttasaari area (Helsinki), the company employed seven people, including chief development officer, chief technology officer, marketing analyst and a team of developers. Even though the documentation and financial operations of the company have been carried out in Finnish language, Ferus Bestia Oy established English as the official working language of the company. That can be easily explained, taking into account that foreign employees constituted the most part of the company. Additionally, partially oriented toward the domestic market, the firm aimed to enter international markets as well.

The primary focus of Ferus Bestia Oy has been set on technology-based information security solutions. For the first security application, Lioncrypt, the development team of the Ferus Bestia Oy attempted to utilize the top technology trends. A number of up-to-date technologies were adopted, this means that topical issue of usability could not remain abandoned. However, the usability evaluation study was postponed, because it normally requires a decent amount of time and effort.

Speaking of Ferus Bestia Oy financial foundation, it has been mainly provided by Tekes – Finnish Funding Agency for Technology and Innovation. Also, valuable contributions were made on behalf of individual investors. Unfortunately, for the year 2016 Ferus Bestia Oy suspended business activities due to a lack of investment. Nevertheless, current research development and implementation were performed in the prior period. Hence, the results of the research are valid, and the recommendations can make use in the future.

2.2 Product description and intended users

Specializing in providing security solutions, the company focused on the development of *Lioncrypt,* software aimed to protect the privacy of the data stored in various cloud services or mass storage devices.

For the year 2015, Lioncrypt software was comprised of 2 *user interfaces*: desktop application and website, both Lioncrypt software components adopted the English language*.* At its core*, user interface (UI)* enables the interaction between the user and the software.

The working version Lioncrypt 1.0 offered the possibility to protect data stored in Dropbox free cloud storage service. On the whole, Lioncrypt 1.0 can be described as MVP before release. *MVP* or *minimum viable product* – is an early version of product, website or software, which has a minimum set of essential features ready to be exposed to the target customers (Reis 03 August 2009). This approach can provide certain benefits, for example, obtaining the early feedback from users, hence, adjust various aspects before the final design (Tho 2015).

Lioncrypt 1.0 provides two types of user roles available to the customer: admin and user. Admin plays an only administrative role and has a right to add or delete users. All actions for the admin role are performed through the website. As for user, once the desktop application is installed, it is possible to copy files to the application, encrypt files by adding them to the application and store them in Dropbox cloud service. Plus, the option of sharing files between registered Lioncrypt 1.0 users is provided.

At first, the novice user needs to log into the website, add his Dropbox cloud service credentials, download and install the desktop application. After the user logs into the application, he (or she) needs to create Lioncrypt folder where the encrypted files can be seen. It is important to understand that Lioncrypt software *does not store the files in* own servers of the company, but only encrypt them in order to protect the sensitive data stored somewhere else. In case of using Lioncrypt 1.0, the files are being stored in the Dropbox cloud service.

Concerning intended users, they are represented by micro to small size company personnel, security consultants, team leaders and, in general, by everyone concerned about secure data storage.

3 Terms and definitions, related to usability and usability evaluation

This chapter provides the background of the research by focusing on terms and definitions of main usability concepts. At first, the difference between terms usability and user experience is explained. In the second part of this chapter, alternative usability evaluation methods are described and compared.

3.1 Usability and user experience

Over the last decade, usability was often contrasted with user experience. Both are indivisible from the user and the product, however there are subtle differences between them.
This chapter aims to define terms usability and user experience and explain how these terms are related.

3.1.1 Usability definition

At this moment, usability is recognised to be a significant factor defining the success of software, website, product or service (Baguma, Kiprono & Kirui 2016, 46; Dubey & Saxena 2013, 48; Hayat, Lock & Murray 2015; Nielsen 2012d; Pratas 2014, chapter 3).

Often simplified to the ease of use (Jeng 2005, 3), the complex concept of usability can be defined in multiple ways. For instance, Albert and Tullis (2013, 5) narrowed down the usability definition to the successful completion of a task by the user. Whereas, Bevan, Carter and Harker (2015) insist that usability should be considered as a “high-level concept” and offer a multifaceted approach to the notion. In pursuit of better understanding of the subject, numerous experts take another angle on defining usability and refer to it as a quality factor (Bevan 2009b, 107; Moreno & Yague 2012, 168; Nielsen 2012d).

As can be seen, practitioners could not reach consensus on the universal definition of usability. To this end, many interpretations of usability are available. Fortunately, one of the most comprehensive ones comes from International Organisation of Standardisation.

The standard definition of usability, which highlights the crucial usability measures is provided by ISO (9241-11 1998; IEC 25063 2014) “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and

satisfaction in a specified context of use." This the most widely accepted definition specifies the major dimensions of usability, which are explained below.

Effectiveness concerns how accurately and completely users can reach the specific goals (Barnum 2011, 11; ISO 9241-11 1998; Quesenbery 2004, 5). It can be measured, e.g., through task completion rate, the number of persistent errors, the number of tasks, completed with or without assistance (NIST 2007, 19; Travis 2003).

Efficiency indicates the resources spent to accomplish the goals, typically refers to the quickness (Barnum 2011, 11; ISO 9241-11 1998; Quesenbery 2004, 5; Rubin & Chisnell 2008, 4). It is normally measured through time-on-task or time-on-task in relation with task success (Albert & Tullis 2013, 86; Travis 2003).

Satisfaction reflects feelings and opinions of the users about their interaction with the system, product or service (Barnum 2011, 12; Rubin & Chisnell 2008, 4). It includes subjective responses, such as freedom from discomfort (ISO 9241-11 1998) and pleasure received from using the product (Nielsen 2012d). Satisfaction can be measured, for example, through various questionnaires (NIST 2007, 19).

Context of use encompasses intended users, their goals, physical, social and technical environment (ISO 9241-11 1998). In other words, it represents real-world conditions under which system, product or service is used.

Aside from the listed above core attributes of usability, different practitioners proposed to adopt other components. Some of these components expand existing usability attributes, while others serve as independent dimensions. Among a variation of additionally proposed usability attributes, it is worth mentioning learnability.

Learnability defines how much the time and effort put the (novice) user to accomplish the set of tasks (Nielsen 2012d, Quesenbery 2004, 5; Petrie & Bevan 2009). This usability facet demonstrates how fast and easy the user can learn how to interact with the system or product. Rubin and Chisnell (2008, 4) acknowledge learnability as usability aspect, however they determine learnability as a part of the effectiveness. While Grossman, Fitzmaurice and Attar (2009, 649) also recognize learnability as the usability criteria, they emphasize that "there is little agreement as to how learnability should be defined, measured, and evaluated."

Although, Sauro (2013a) tends to consider learnability as a separate dimension from usability, he also offers some practical advice how to assess learnability. According to Sauro (2013a), there are two kinds of learnability: one is related to novice user, another one is related to learnability over time. Learnability over time can be measured via series of trials through comparing the task performance (Sauro 2013a). While the learnability, connected to novice user might be assessed through SUS questionnaire (see section 4.5.2).

As noted earlier, the importance of usability is globally accepted. Notwithstanding this overwhelming evidence, literature research indicates a lack in absolute definition of usability. Moreover, different sets of usability attributes are presented and unclearly explained. Even though the prior research has shown the differences in terminology, it is possible to distinguish the common factors that affect usability: users, goals, effectiveness, efficiency, satisfaction and particular environments.

3.1.2 User experience definition

The definition of usability, explained in the previous subchapter, still remains unfinished and open for the further discussion. As for user experience, it appears to be even more controversial topic.

The International Organization for Standardization (ISO) has been representing series of standards over the years, offering various definitions of user experience. At the moment, user experience can be described as: “person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service” (ISO/IEC 25063 2014). Aiming to improve and clarify this formal definition, additional notes has also been presented in the standard paper. In a nutshell, the notes attempted to reflect all possible aspects of user experience, e. g., emotions, beliefs, context of use, skills and prior experiences of the user, his personality. (ISO/IEC 25063 2014).

Nielsen & Norman suggest a more simplified definition of user experience, which basically covers everything related to the ”end-user’s interaction with the company, its services, and its products.” Nevertheless, Bevan et al. (2015, 152) argue that this interpretation perhaps just leads to further confusion.

Overall, the literature review illustrates that the definition of user experience is still under debate. However, the majority of practitioners agree on the significant facet in user experience: subjective features, such as joy-of-use and emotional experience

(Hassenzahl 2008; Rauschenberger, Schrepp, Perez-Cota, Olschner & Thomaschewski 2013, 39).

3.1.3 Usability vs. user experience

The discussion above shows that usability and user experience are not given precise definitions. Frequently, the same terms are used to characterize different concepts. Furthermore, the scope of these two concepts overlap.

Regardless of the implications, multiple practitioners recommend distinguishing user experience and usability (Bevan & al. 2015; Guo 2012; Morville 2004; Nielsen & Norman; Petrie & Bevan 2009).

Various types of research tagged with analyzing the relationship between usability and user experience have been done. A number of experts consider usability as a part of user experience (Guo 2012; Morville 2004). Figure 1 demonstrates well-known user experience honeycomb, developed by Morville (2004).

Figure 1. User Experience Honeycomb (Morville 2004)

In this case, “usable” hexagon refers to usability, which represents ease of use (Morville 2004). Other facets of the honeycomb relate to the concept of user experience and are out of scope of this thesis. The same logic underlies another illustration of the relationship between usability and user experience.

Value	Usabilty
Is it useful?	Is it easy to use?
Adoptability	**Desirability**
Is it easy to stop using?	Is it fun and engaging?

Figure 2. Four elements of user experience (Guo 2012)

It can, therefore, be assumed that user experience comprises the combination of multiple components, including usability. This notion is also supported by Robier (22 March 2015), who incorporates various processes taking place before, during and after the interaction with the product or service.

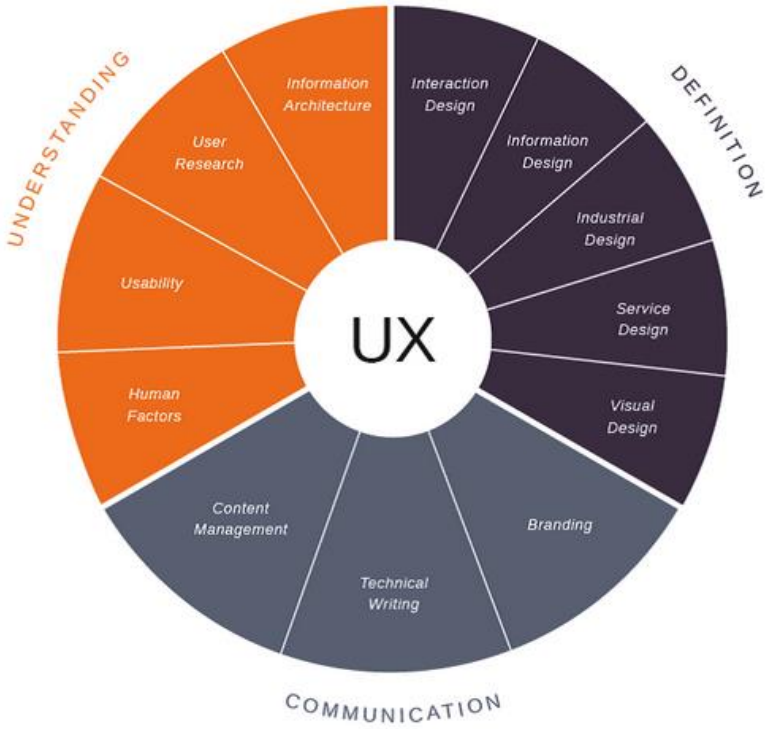


Figure 3. User experience Robier (22 March 2015)

To conclude, the difference between usability and user experience remains a topic open to interpretation. User experience represents a broader concept than usability; it might include multiple factors, such as, e.g., branding, marketing, design, value, aesthetics or emotional factors.

In light of the above, it appears obvious that the user experience concept goes beyond the scope of this thesis, accordingly, from now on the author will concentrate on the prime focus of this study – usability.

3.2 Usability evaluation

Nowadays, a number of types of usability measures have been developed to facilitate a usability evaluation. Even though the classification differs greatly, these methods pursue the certain aims: acquire the more accurate knowledge about the needs of users, contribute to product improvement and increase the value for users by achieving the higher quality user experience (Bevan 2008). A usability evaluation, no matter which method is to be adopted, can positively affect team decision-making by relying on sources distinct from personal judgments (Schade 2013b).

To assess the usability, multiple usability evaluation procedures were gathered under one roof. Although the categorization might vary usability evaluation methods can split broadly into two categories, depending on the involvement of typical end-users into assessment process: analytical and empirical. Usability inspection depicts analytical methods, where usability experts inspect an artifact. On the other hand, usability testing represents empirical methods as actual users or their representatives interact with the system. (Cocton, Woolrych & Lavery 2007, 1172; Plantak, Kirinic & Klicek 2010, 272; Sauro 2011c.)

3.2.1 Usability inspection

Usability inspection approach includes a variety of methods, which are based on examining interfaces by usability specialists, instead of active user participation (Hollingsed & Novick, 2007, 2; Nielsen 1995a). Two frequently used usability inspection methods are heuristic evaluation and cognitive walkthrough (Davids, Chikte & Halperin 2013, 242).

Heuristic evaluation is one of the most widely applied usability inspection methods (Wilson 2014, chapter 1). This method is conducted by usability specialists evaluating user interface against predefined set of rules (Wilson 2014, chapter 1).
Typically, heuristic evaluation method is based on ten guidelines or so-called “heuristics”, developed by Nielsen (1995b), for example, consistency and standards, error prevention, help and documentation and so on. Multiple practitioners ordinarily use these heuristics, however, many researchers prefer to extend Nielsen’s techniques (Georgsson, Staggers & Weir 2016, 78) or select their evaluation principles (Shneiderman 2010).

Heuristic evaluation is preferably applied on early stages of the development process (Davids & al. 2013, 248; Zazelenchuk 2006), starting from the point when any kind of user

interface is present, with the aim of identifying design flaws. Besides, heuristic evaluations also make use at any phase of the development stages (Wilson 2014, chapter 1) to determine whether UI design follows the established set of principles.

A cognitive walkthrough is another commonly adopted usability inspection method. Usually performed by one or more usability evaluators, this method has been employed to estimate how well the actual user would carry out the tasks toward the selected goals (Kolski, Mahatody & Sagar 2007, 1; Wilson 2014, chapter 4). As a rule, cognitive walkthrough includes the questioning part accompanying the predefined set of tasks to examine the learnability of each activity (Lewis, Polson, Rieman & Wharton 1994, 9; Svee & Zdravkovic 2014, 2). Various alterations were offered to improve cognitive walkthrough usability inspection method; it can be modified, extended or use combined with other usability evaluation techniques (Kolski & al. 2007; Svee & Zdravkovic 2014). Similar to heuristic evaluation method, a cognitive walkthrough can be applied at any stage of the development process.

As discussed above, usability inspection methods often do not require end-user engagement or somewhat minimize it. Inspections are ordinarily performed by usability professionals or evaluators available. Usability inspection methods can be used separately or in combination with another usability evaluation methods, frequently applied in the early stages of the development process, but most likely remain useful during other stages.

3.2.2 Usability testing

Usability testing definitions vary from the simplified ones, which state that usability testing is nothing but the capability of an ordinary person to find out how to perform actions smoothly and efficiently while using the artifact (Krug 2014, 9), to more complex clarifications.

For instance, Rubin and Chisnell (2008, 21) describe usability testing as follows: ”Usability testing – process that employs people as testing participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria”. Barnum (2011, 13) explains the core meaning of usability testing briefly: ”Usability testing – activity that focuses on observing users working with a product, performing tasks that are real and meaningful to them.” Yet, she admits that the given main definition can expand, depending on the scope of testing, environment, methods and other conditions (Barnum 2011, 14). In summary, usability testing can be defined as a

method, tool or process, during which representatives from real users are asked to perform a set of tasks in an attempt to discover usability issues and improve the product (Barnum 2011, 6; Rubin & Chisnell 2008, 21).

Usability testing is sometimes conducted only during the later stages of the development. However, a number of experts recommend to perform usability testing throughout the development lifecycle, starting from the early stages (Petrie & Bevan 2009, 22; Rubin & Chisnell 2008, 25). Moreover, usability testing provides the most value, when it is an ongoing iterative process (Krug 2014, 124; Petrie & Bevan 2009, 23; Rubin & Chisnell 2008, 28). This approach enables to detect and fix usability issues as soon as they appear, thus, improve overall usability (Sauro 2014a).

It is a common practice to create a *scenario* for usability testing, based on a set of tasks, through which the recruited participants attempt to communicate with UI, as they would act in reality (Barnum 2011, 19). Probably, the main benefit of using scenarios is that it enables to find out how users de facto communicate with software or product (Sauro 2013c).

Different types, methods, and techniques are gathered up under the expanding umbrella of usability testing. Chapter 4 of the present thesis offers an overview of well-known methodologies; especially detailed coverage is provided to the methods closely related to the study.

3.2.3 Summary of usability evaluation methods

Aiming to improve usability, researchers developed a tremendous number of approaches. Analytical and empirical methods have been widely employed for evaluating usability. Usability inspection methods are classified as analytical, handled by the experts and usually eliminate or minimize the participation of the actual users. On the other side of the coin, empirical usability testing normally involve real users to capture their interaction with the system.

Certainly, both approaches have advantages and weaknesses. Usability inspection methods are traditionally considered reliable, time and cost saving (Davids & al. 2013, 243; Cocton & al. 2007, 274). They are easier to implement in early stages of development process than usability testing, especially while evaluating prototype (Cocton & al. 2007, 274; Sauro 2011c). Nonetheless, in order to perform valid and effective usability inspection, usability professionals should be involved in the process. Moreover, it

requires the participation of multiple experts (Cocton & al. 2007, 243; Macefield 2014; Sauro 2011c). In this context, researchers might face a range of challenges, including increasing costs and inability to find a sufficient number of usability experts.

Another key fact to remember, usability inspection methods highly depend on skills, knowledge and experience of assessors (Ardito, Costabile, De Angeli & Lanzilotti 2006, 196; Matera, Rizzo, & Carughi 2006, 159), that probably makes this kind of approach less objective than usability testing.

Usability testing is widely reckoned to be the most efficient way to evaluate usability due to the direct involvement of real users (Ardito & al 2006, 196; Matera & al. 2006, 159). As stated by Sauro (2014b), “in an ideal world, users would be involved in every stage of product development.” Furthermore, he claimed that there is no real alternative to usability testing, as for other usability evaluation methods – they might be a suitable addition to it (Sauro 2014b). Another compelling argument is that usability testing facilitates to persuade developers and stakeholders that certain usability issues are present in the UI, as the feedback is coming from the actual users (Perfetti 2003).

While usability testing allows revealing usability issues under the conditions similar to the real world, it also has certain limitations. Among the major disadvantages of usability testing are time, cost, difficulty to find eligible participants, struggle to test on low-fidelity prototypes (Davids & al. 2013, 243; Rubin & Chisnell 2008, 26; Sauro 2014a).

Given these points, numerous practitioners supported the idea of combining distinct methods as it might combine their strength as well (Barnum 2011, 71; Davids & al. 2013, 248; Cocton & al. 2007, 1172; Hollingsed & Novick 2007, 252; Sauro 2014b). For instance, famous usability experts Barnum (2011, 71) and Sauro (2011c) recommend to couple usability testing and heuristic evaluation techniques. Nevertheless, as has been mentioned, heuristic evaluation and other analytical methods unable to substitute entirely for usability testing (Hollingsed & Novick 2007, 252; Sauro 2014b).

All things considered, it has to be concluded that analytical approach cannot substitute empirical usability evaluation methods. However, it can become a complemental addition to the main method. Under the limited resources condition and in order to get the best results, it is preferable to choose empirical usability testing over usability inspection methods. In case if usability testing option is unavailable, it would be advisable to put usability inspection method(s) into practice. After all, it is an acknowledged fact that it is better to test, than not to test at all.

Generally speaking, usability evaluation consists of a wide range of methods and has become a crucial part of software development process. As a matter of fact, usability professionals have not established a uniform standard for usability evaluation methods; there is no universal approach for each and every case. On one hand it could create difficulty in choosing the right method, on the other hand, increase the amount of flexibility, make room for multiple techniques integration and facilitate adjusting usability evaluation strategy to meet the particular goals.

4 Methodological approaches to usability testing

This chapter focuses on the widely adopted methods and techniques for empirical usability testing. Firstly, the pros and cons of various forms of usability testing are compared and contrasted, then methods and techniques employed into current study are described in detail. Finally, practices of recruitment and selection of the participants are discussed.

4.1 Usability testing types: study goals perspective

Usability testing comes in a variety of flavors. In order to choose one method over another, it is important to take into account multiple factors, study goals and product development stage take a respectable place among them. Generally, usability testing can be categorized as formative and summative.

Formative usability testing is aiming to diagnose usability problems to make necessary improvements. It is usually performed throughout development cycle in an iterative manner. (Albert & Tullis 2013, 42; Barnum 2011, 14; Rubin & Chisnell 2008, 28.)
Due to the main goals of formative testing (identifying usability issues and gaining insight into end users behaviors), it is sometimes called *exploratory* (Dumas & Fox 2009, 236; Rubin & Chisnell 2008, 27). Besides, Sauro (2015b) use the term *problem discovery* to refer to formative usability testing.

The term s*ummative usability testing* usually implies final assessment or evaluation, conducted during the late-stage development. The goal of this kind of testing is to determine how well product or service meets a set of requirements. (Albert & Tullis 2013, 43; Barnum 2011, 14; Sauro & Lewis 2012, 10.)

Despite different goals, formative and summative types of usability testing usually employ similar methods, focusing on different study details. Formative usability testing often pertains to small studies, whereas summative usability testing frequently entails a large number of subjects (Barnum 2011, 14). Apart from that, formative studies concentrate on insights, while summative are targeting statistical data.

Accordingly, formative studies are typically *qualitative (non-numerical)* by nature, while summative research is *quantitative* (numerical). (Hodgson 2010.) Nonetheless, regardless of the chosen usability testing type, collecting *both qualitative and quantitative data* may result in deeper coverage of usability issues (Hodgson 2010; Sauro & Lewis 2012, 10).

4.2 Usability testing types: location perspective

This section describes and contrast alternative usability testing types from the location point of view. Generally, usability testing can be classified into two main categories: traditional in-lab usability testing and remote usability testing.

4.2.1 Traditional in-lab usability testing

Traditional usability testing or in-lab usability testing is the most frequently applied method (Albert & Tullis 2013, 53). As a rule, traditional usability test involves the representatives of the real or potential users, asked to carry out the series of realistic tasks while being observed in usability lab (Albert & Tullis 2013, 53; Perfetti 2010). Regularly, the only one participant goes throughout the scenario (or set of tasks) during one test session, guided by the so-called *moderator* (Albert & Tullis 2013, 53).

Rubin and Chisnell (2008, 45) describe the moderator as “the one team member that you absolutely must have in order to conduct the test”. Spool (2009a) compares moderator with an “orchestra conductor” and Sauro (2015b) refers to the entire in-lab usability testing approach as “moderated in-person”. *The moderator* is the person which responsibilities include, for example, preparation of the test materials, interaction with the participant, monitoring his or her behavior, overall test administration, and data collection. Also, the moderator should stay objective and do not over interfere in actions performed by the participant. Given the all the responsibilities assigned to the role of moderator, it comes as no surprise that usability experts considered this role critical and immensely challenging. (Barnum 2011, 162; Rubin & Chisnell 2008, 45.)

Apart from the moderator, depending on the methods and techniques chosen for the particular research, traditional in-lab usability testing might engage observers, loggers, note takers, technicians and even help desk operator (Barnum 2011, 163). Evidently, in-lab usability testing provides excellent conditions for the observation technique (Fluckiger & Richter 2014, 65), considered later in this chapter.

Speaking of usability labs, they exist in a large variety of types (Barnum 2011, 34). Initially, in-lab usability testing was associated with state-of-the-art usability laboratory, including one-way mirror, several rooms, and sophisticated equipment; obviously, it

resulted in excessively expensive activity (Krug 2014, 115). However, usability practitioners were looking for a way to conduct usability tests in a cost-effective manner.

Nowadays, numerous usability professionals believe that it is possible to execute valid usability tests in the *informal lab*, for example, conference room or office (Barnum 2011, 26; Krug 2014, 122; Nielsen 2012c; Travis 2013a). It can, therefore, be assumed that the magnificent opportunity to perform efficient and reliable testing while staying on budget is just around the corner.

Typically, in-lab usability testing is utilized in the formative studies for catching and fixing defects during iterative design improvement process. Frequently, this kind of testing involves a small number of respondents and is coupled with thinking aloud technique, discussed later in this chapter.

4.2.2 Remote usability testing

Remote usability testing has a similar structure to traditional usability testing. However the critical difference is that during the remote usability testing participants and researchers are located separately (Barnum 2011, 42; Fidas, Katsanos, Papachristos, Tselios & Avouris 2007, 152; Schade 2013a).

The possibility to reach out the users from around the world is the key aspect of the remote usability testing (Baker 2014). The test session might take place in the lab or even within an environment which is natural for the participant, for example, work or home (Tullis, Fleischman, McNulty, Cianchette & Bergel 2002). Remote usability testing can be organized into two categories: moderated and unmoderated (Baker 2014; Travis 2007).

Moderated remote usability testing requires the involvement of the moderator, who instructs and guide the participant remotely throughout the test session (Baker 2014; Barnum 2011, 41; Schade 2013a). This approach is also called synchronous because the data is being collected in the real time, although the facilitator and the participant are physically separated (Barnum 2011, 42).

During *unmoderated* remote usability testing, respondents accomplish predetermined set of tasks without a moderator present (Albert & Tullis 2013, 54; Baker 2014). Unmoderated testing is also known as *automated* due to the reason that the data is presented and collected via software tool (Barnum 2011, 44; Soucy 2010). In contrast with moderated remote usability testing, this kind of test does not require real-time human interaction: the

data recorded during test sessions is examined later by the usability professionals (Schade 2013a). This *asynchronous* approach allows the participants to complete test sessions concurrently and hence, generally test large volumes of participants (Albert & Tullis 2013; Baker 2014).

Each of the remote usability testing variations has its pluses and minuses. Remote moderated usability testing gives more flexibility as a moderator can alter the process. Hence, it allows better task control and helps to receive insightful data. In the same time, remote unmoderated usability testing grants the possibility to test hundreds of participants simultaneously (Soucy 2010).

4.2.3 Traditional vs. remote usability testing

As can be seen from the preceding discussion, traditional in-lab usability testing and remote usability testing are both valid and mature methods for usability evaluation. For the moment, there is no consensus among researchers regarding the best way to perform usability testing. Despite this fact, one or another testing kind might be chosen depending on the study goals, context, and limitations.

For example, remote usability testing might be suitable when there is a need to test participants from across the globe or it is difficult to schedule an in-lab test. Besides, unmoderated remote usability testing is an effective way to obtain significant amount of feedback. However, the weaknesses of the remote usability testing include the fact that it heavily depends upon third-party services, such as complicated software or Internet connection. Participants might face miscellaneous technical problems or struggle to set up tools intended to be used for performing remote test sessions (Barnum 2011, 338; Bolt 2010; Travis 2013a). Also, collecting visual feedback from the participants (e. g., body language, facial expressions) can be challenging and is rarely employed during remote usability testing (Barnum 2011, 338; Brierley 2014; Fidas & al. 2007, 153). These arguments suggest that traditional usability testing offers better control over the testing environment and other conditions.

Another factor to consider when selecting a type of testing is cost. Today, there is an ongoing discussion about the cost of remote and traditional usability testing. As noted earlier, traditional usability testing used to be associated with higher costs. However, cost-effective methods generate accurate results as well (Krug 2014, 116; Nielsen 2009). Undoubtedly, these days traditional in-lab usability testing comes at a much smaller price tag. On the other hand, remote usability testing, often considered as an inexpensive

testing alternative (Bolt 2010; Sauro 2012a; Tullis & al. 2002), can be even more expensive than in-lab testing, because of the high prices for the testing software.

Also, the choice of usability testing depends on the formative or summative nature of the study. As mentioned previously, traditional usability testing is typically used in the formative studies, whereas remote usability testing is frequently employed in summative studies.

Above all, the majority of usability practitioners reported that remote moderated usability testing and traditional in-lab usability testing showed similar results (Andreasen, Nielsen, Schroder & Stage 2007, 1413; Thompson, Rozanski & Haake 2004, 136; Tullis & al. 2002). That means these approaches could be described as equivalent alternatives for usability evaluation study.

Even so, some researchers still find traditional in-lab testing more reliable than remote usability testing, especially comparing to its unmoderated variation (Brierley 2014; Travis 2014). The idea that traditional usability testing most likely provides richer user insights and could facilitate to uncover specific usability issues (Cerejo 2016) was more or less supported by the study carried out by Tullis & al. (2002). Although the study results revealed that both approaches brought to light usability problems in the similar degree, certain types of user behaviors were captured only during the in-lab test session.

Finally, it would be appropriate to quote well-known user experience professional David Travis (2013a): “If the aim is to expose the design team to real user behavior, then not much will beat a test in a corporate lab or a rented facility”.

4.3 Think-aloud protocol

The think-aloud protocol, also called thinking aloud or verbal protocol, is a method involving users to vocalize their thoughts while they are performing the required tasks (Bergstrom & Olmsted-Hawala 2012, 86; Chisnell & Rubin 2008, 204; Nielsen 2012b).

The think aloud protocol is commonly used by usability practitioners. Furthermore, Nielsen (2012b) calls thinking aloud the number one usability testing method. The popularity of this method comes from a variety of advantages, such as simplicity, low-cost implementation, and flexibility. Essentially, as stated by Nielsen (2012b) “it serves as a window on the soul”, which means that researchers gain an opportunity to reveal what users actually keep in mind while interacting with the system, product or service.

Probably, the main limitation of the think-aloud protocol is that it can create an obstacle for non-native English speaking participants if the study performed in English. However, a comprehensive view of this issue cannot be formed due to lack of research.

Thinking aloud can be conducted in two general ways: concurrently and retrospectively. During the most common form of think-aloud protocol – concurrent think aloud (CTA) participants complete tasks and describe their thoughts, questions, and feelings in parallel. (Rubin & Chisnell 2008, 204; Bergstrom & Olmsted-Hawala 2012, 86). Since thinking aloud is an unnatural behavior for the majority of people, the participants might fall silent from time to time (Isbister & Schaffer 2008, 70; Nielsen 2012b). In this case, they should be neutrally encouraged to keep talking (Rubin & Chisnell 2008, 205; Isbister & Schaffer 2008, 70).

An alternative approach to thinking aloud is called retrospective think aloud or RTA. This method involves participants interacting silently with an artifact and report their thoughts after the performance is over (Bergstrom & Olmsted-Hawala 2012, 86; Elling, Lentz. & Jong 2011, 1161). Ordinarily, using audio/video recording or logging facilitate more accurate recall of their experience (Haak et al. 2003, 341).

Comparing these two methods of think-aloud protocol, different strengths and weaknesses can be indicated. Probably, the main concern of CTA method is that it can affect performance and increase the time taken to complete tasks (Haak, Jong & Schellens 2003, 349). Whereas more recent study carried by McDonald & Petrie (2013, 2943), concluded that concurrent thinking aloud has no influence on performance, however, increases frustration and effort on tasks.

In the same time, RTA can considerably prolong the duration of the testing session, for the reason that participants primarily complete the tasks, and then review them in retrospect (Isbister & Schaffer 2008, 68; Rubin & Chisnell 2008, 55). Besides, the delay in reporting can cause forgetting information by participants or lead to reconstruction, i.e. different interpretation of performed actions (Haak et al. 2003, 341; Rubin & Chisnell 2008, 55). In this respect, concurrent think aloud protocol seems more reliable, then its retrospective alternative. Nevertheless, researchers generally express the opinion that CTA and RTA achieve similar results and can be considered as analogous methods of thinking aloud (Bergstrom & Olmsted-Hawala 2012, 92; Haak et al. 2003, 350).

All in all, think-aloud protocol assist developers in understanding the way of thinking of end-users. Even though collecting and analyzing verbalizations might be time-consuming (Isbister & Schaffer 2008, 69), it allows to obtain meaningful and useful quotes (Holzinger 2005, 73). Apart from that, the think-aloud protocol is a marvelous way to convince developers and designers on the grounds that feedback comes directly from end-users (Nielsen 2012b).

4.4 Observation techniques

Observation techniques enable to collect data while watching participants interacting with the product, software or service. The data collected might include, e.g., facial expressions, body language, gestures or sighing. (Barnum 2011, 138.) It is recommended to invite developers and stakeholders to observe test sessions because it establishes a higher level of trust in study findings (Nielsen 2010). *Observers* stay silent during a testing session and normally take notes (Barnum 2011, 227).

Aside from taking notes, *logging* technique can facilitate to record verbal and non-verbal observations. The role of *logger* includes monitoring the test session, collecting the data about user interaction with UI and note it down by using specific codes. In addition, the logger might add a short description of non-verbal observations, quotes or other appealing expressions. (Barnum 2011, 225; Travis 2010.) In practice, logging can be done manually or using various data logging software (Zazelenchuk 2008).

Commonly, observation is supported by capturing on-screen activity, video, and audio recording. These techniques are utilized for the purpose of further review and analysis. Summing up, observation encompasses powerful techniques which are linked with usability testing, as stated by Sauro (2015c): “Observing just a few participants attempt actual tasks can reveal a great deal about interaction problems and generate ideas on what to fix.”

4.5 Questionnaires

Questionnaires facilitate to gather feedback from participants before, during and after the usability testing (Barnum 2011, 173). Pre-test questionnaires aim to collect information about the background of the participant, post-task questionnaires help to capture instant reactions, and post-test questionnaires allow to aggregate the responses about overall experience (Dumas & Reddish 1999, 209).

Typically, questionnaires can be open-ended and closed-ended. Open-ended questionnaires provide to a participant possibility to answer in own words, while closed-ended questionnaires limit the answer choices to the fixed set of alternatives (Sauro & Lewis 2012, 185). This subchapter provides details regarding popular questionnaires, which were used in current research.

4.5.1 Screening questionnaire

The purpose of a screening questionnaire is to select the appropriate test participants for a study (Rubin & Chisnell 2011, 126). A decent screening questionnaire should echo the specific characteristics outlined for the target group of users (Nielsen & Sova 2003, 29).

As stated by Dumas & Reddish (1999, 143), a screening questionnaire has two main objectives:

- rapidly identify if the subject is an acceptable participant for the test
- define to which subset of the target audience the potential participant belongs.

To put it briefly, a screening questionnaire is normally sent to the potential participants before the study to ensure that they match the defined criteria for the study. After that selected suitable participants are recruited for the usability testing.

4.5.2 System Usability Scale (SUS)

Originally introduced by Brooke in the mid-eighties, the System Usability Scale (SUS) has de-facto become an industry standard by virtue of its simplicity and free availability. Among multiple benefits gained from using SUS, it is important to note that it can be applied basically to any sample size, even to very small one, without affecting the accuracy of results. (Brooke 2013, 29, 38; Sauro 2011b.)

The System Usability Scale (SUS) is a ten-statement questionnaire widely applied for the usability assessment of miscellaneous types of user interfaces (Albert & Tullis 2013, 137; Bangor, Kortum, & Miller 2009, 115; Brook 2013, 29). Each of these ten questions is accompanied by a five-point scale of agreement, providing the selection of response options from “strongly agree” to “strongly disagree.”

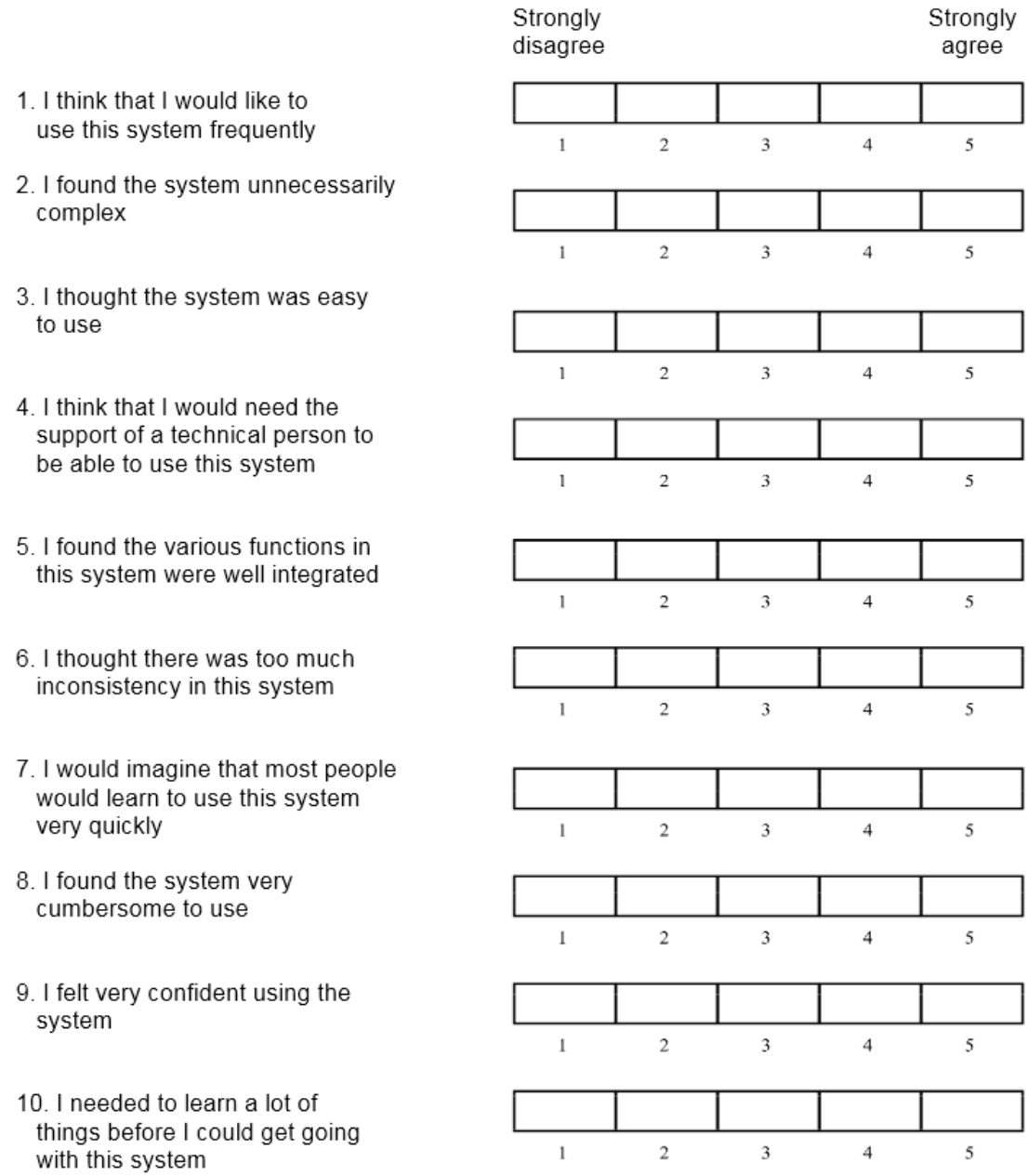


Figure 4. SUS sample (Brooke 1996, 192)

As seen from the figure 4, half of the statements are formulated negatively and another half positively. This approach enables to reduce response bias through encouraging the participants to answer the questions thoughtfully (Brooke 1996, 191).

Later on, Sauro (2011a) questioned the necessity of including both positively and negatively worded items into SUS and offered “all-positive” version of SUS questionnaire. Brooke (2013, 34) admits the possible benefits of an alternate version. However, he advocates the traditional version of SUS and points out its popularity and credibility. Some other usability experts, including Barnum (2011, 182) endorsed the classic version of SUS questionnaire as well.

Next, a valuable adjustment to the common version of SUS questionnaire has been put into practice by quite a few usability specialists, for example, Finstad (2006), Bangor & al. (2009), Kortum & Peres (2015). They recommended replacing the word “cumbersome” with the word “awkward” in statement 8, especially for reseaches engaging non-native English speakers. Furthermore, in several studies the word “system” has been substituted with other research-related words, for example, “product” (Bangor & al. 2009, 115), “website” (Sauro 2011a) or “medical device” (Kortum & Peres 2015). Even though minor alterations enumerated above do not change the results, it is a good practice to avoid excessive modifications to protect the validity of original SUS questionnaire (Barnum 2011, 182).

The SUS results are converted into a single score, which ranges from 0 to 100, where the higher numbers indicate better usability. The overall score calculations are scrupulously explained in the literature. Additionally, ready-made calculating tools are widely available on the Internet (Albert & Tullis 2013, 138; Barnum 2011, 182; Sauro 2011b).

Even though the SUS questionnaire is frequently used, there is a lack of guidance about the interpretation of the results (Sauro 2011b). One option would be to follow the recommendations of Bangor & al. (2009, 121) and interpret the SUS score according to the following scale.

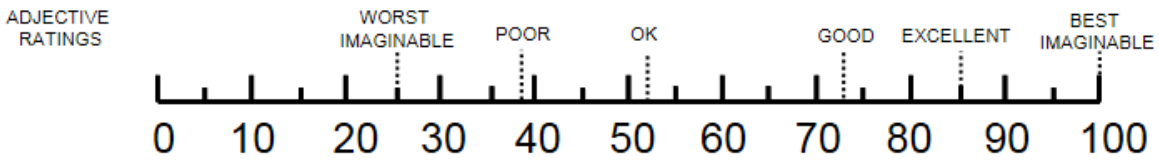


Figure 5. Interpretation of SUS score according to Bangor & al. (2009, 121)

Besides, (Sauro 2011b) offers to compare the SUS score to the average SUS score from 500 studies. In this case, the SUS score below 68 would be below average while the score above 68 is above average.

In addition, Lewis & Sauro (2009) reported that it is possible to assess *learnability* through the SUS questionnaire. They defined two dimensions: learnability (calculated by multiplying the sum SUS contribution scores from the items 4 and 10 by 12, 5) and usability (the sum of other items SUS scores, multiplied by 3,125). Sauro (2013d) expresses the opinion that measures of learnability might be of equal importance as usability assessment for some studies.

4.5.3 Single Ease Question (SEQ)

The Single Ease Question (SEQ) is a seven-point scale post-task questionnaire evaluating the user impression given by the task performed (Sauro 2015a, 122). Simple rating scale allows to determine how challenging the task completion was for the user. Similar to SUS questionnaire, participants can rate their experience from “very difficult” to “very easy”. According to the study by Sauro (2012d), who collected the results from 5000 users, the “average score hovers between about 4.8 and 5.1”. The typical example of the SEQ is shown in the figure below.

Figure 6. SEQ sample (Sauro 2015a, 122)

SEQ has a number of benefits: it is short and easy to fill, it can be used across various devices, websites, software applications or paper prototypes, additionally, it is easy to administer and interpret (Sauro 2010; Sauro 2012d). Apart from that, SEQ remains robust even for small sample sizes (Sauro 2012b).

In general, post-task questionnaires allow obtaining instant feedback from participants (Barnum 2011, 176). After the participants completed the tasks one by one, they gain an opportunity to gather their impressions and evaluate their experience. Another advantage of post-task questionnaires is that ”varying the flow of the test during a long session can help to relieve boredom and fatigue” (Rubin & Chisnell 2008, 193).

Even though the notion of SEQ seems to be very basic and incredibly simple, Sauro (2012d) claims that the implementation of SEQ shows as valuable results as more elaborate solutions.

4.5.4 Open-ended question

From the discussion made earlier, it can be concluded that closed-ended questionnaires might be beneficial for usability evaluation, however, they are limited to the choice of preselected statements. Open-end questions, on the other hand, enable to expose personal thoughts and opinions that possibly remained unconcealed (Sauro 2012f).

According to Albert and Tullis (2013, 158), incorporating open-end questions as an addition to other types of techniques has become a common practice in usability studies. The open-ended question might serve as a valuable addition to closed-ended questionnaires. For instance, Barnum (2011, 182) gives the recommendation to include it at the end of the SUS questionnaire to explore peculiar features of the product or service. The formulation of each question may vary depending on the study goals. Most often, open-end questions request answers regarding what user like or dislike or what would participants improve. (Albert, Tedesco & Tullis 2010, 148.)

As can be seen, including the open-ended question in research might be beneficial in terms of receiving the missing data. Also, as stated by Sauro (2012f), that enables to get an insight "what's top of mind for respondents."

4.6 Recruiting and selecting the participants

Participant recruiting for usability testing is an essential component of usability studies (Nielsen & Sova 2003, 8). During the recruiting process, it is important to consider such major issues as a number or participants and general criteria for selection.

According to Nielsen (2000), the best usability testing result might be obtained by recruiting only five users per test session. That allows to discover main usability problems and enables to distribute the budget wisely. Figure 7 displays curve, often cited in the literature, explaining the relationship between a number of test users and usability problems found.

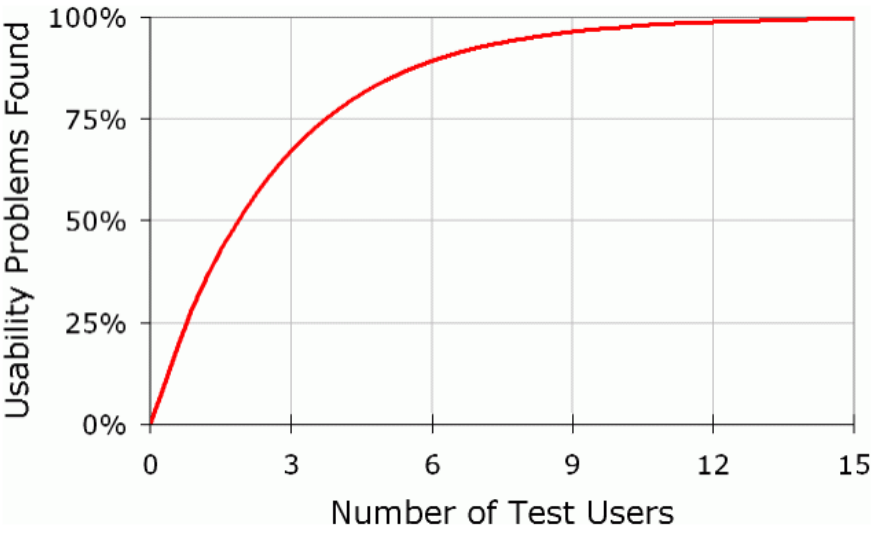


Figure 7. Determining the optimal number of users (Nielsen 2000)

On the other hand, this claim was challenged by Spool and Schroeder, who conducted the study, which demonstrated that five users were able to find only 35% of all usability problems (Spool & Schroeder 2001, 286). Based on collected data, they assumed that five users are most likely not enough. Later on Perfetti and Landersman (2001) carried out the similar study and came to the same conclusion about the number of participants as Spool and Schroeder.

Nevertheless, Nielsen (2012a) argued back, pointing the fact that conducting *iterative* usability testing with five users sample group helps to discover the most critical usability issues. He presented the result of 83 studies and asserted that so-called “the magic number 5” not only works quite well in the most cases, but also enables to achieve the optimal balance between costs and benefits (Nielsen 2012a).

After the years of debates and discussions, there is no consensus on the number of participants. However, many agreed that the final number of participants depends on study goals and homogeneity of the target audience.

For instance, for formative (or problem-discovery) usability testing from five to eight participants is most likely enough (Petrie & Bevan 2009, 23; Six & Macefield 2016; Wiklund, Kendler & Strochlic 2016, 115). To the contrary, summative usability testing usually involves around 30 participants and more (Albert & Tullis 2013, 59; Petrie & Bevan 2009, 23; Wiklund & al. 2016, 115).

Aside from the above, special attention should be given to the differences between users. For the small (formative) studies participants should be chosen from one subset of target audience. Whereas for the large (summative) studies it is recommended to consider different subgroups or increase the size of the sample. (Barnum 2011, 18; Wiklund & al. 2016, 115.)

The selection of participants who represent potential users of the product is an essential element of the testing process. Furthermore, selecting representatives of the target users correctly ensure the validity of the test results. (Albert & Tullis 2013, 58; Rubin & Chisnell 2008, 115). Creating *personas* might well serve as a basis for selecting participants for a usability testing. Each *persona* represents a fictional user from the specific group of the target audience. Personas created based on the preliminary research, describe target users personal characteristics, behavior patterns, and needs. (Rubin & Chisnell 2008, 94; Sauro 2012c).

To sum up, usability professionals have not found a final answer to the question how many participants should be involved in a usability testing. However, it can be recommended to make a choice depending on the usability testing type, goals, and limitations. In a nutshell, for many studies “it’s enough to test 5 users to get a good idea of the main usability insights” (Nielsen 2010). Also, a sampling of the target audience and selecting participants, who share a set of characteristics with intended users, make a large impact on study results.

To conclude, in this chapter types of usability testing, methods and techniques were contrasted and discussed to facilitate the creation of the framework for reliable usability evaluation. The next chapter provides the rationale for the methodology employed in this research and describes the actual experimental procedure.

5 Usability evaluation study on Lioncrypt 1.0

This chapter presents the detailed overview of usability evaluation study on Lioncrypt 1.0. Starting from the outline of objectives and research questions, the framework design and thinking behind the choice of methods is introduced. Environment, criteria for selection of participants and actual procedures are thoroughly described.

5.1 Usability evaluation study questions

The usability evaluation study is designed to address the following questions:

- What is the attitude of the participants towards the use of Lioncrypt 1.0?
- How easy is it for a novice user to learn how to use Lioncrypt 1.0?
- Which scenarios are the most difficult for novice users to accomplish?
- What are the usability problems of Lioncrypt 1.0?
- What are the recommended solutions to fix these problems?

5.2 Framework design and rationale

Based on the approaches, methods and techniques discussed earlier, the framework for usability evaluation of Lioncrypt 1.0 has been developed. This chapter aims to assist in understanding the thinking behind the choice of methods.

Firstly, the major approach for usability evaluation was determined. The discussion, which took place in chapter 3 resulted in the conclusion that *usability testing* (empirical approach) is the most efficient way for the usability evaluation of the Lioncrypt 1.0. To recap, this approach serves to provide findings, derived directly from the ideas and behaviors of the actual users (or their representatives). In contrast, analytical approach lacks the realism, because the evaluation is usually performed by usability experts or via involving hypothetical users (see chapter 3). Besides, as has been mentioned earlier, usability testing is perhaps the best method to convince developers to take into consideration usability problems, as this comments are coming from real users

Secondly, according to the *usability evaluation objectives*, outlined at the beginning of this chapter, and the *development stage* of Lioncrypt 1.0 (discussed in chapter 2) the usability testing approach used in this thesis can be described as *formative*. To recap, formative usability testing aims to diagnose usability issues and fix them before release, as stated by Albert and Tullis (2013, 42).

Next, *traditional in-lab usability testing* is recommended for formative studies (see section 4.2.1). Also, taking into account the limited sources of a start-up company, traditional usability testing appeared to be the best choice it terms of budget and amount of the participants. The steps listed above are demonstrated in figure 8.

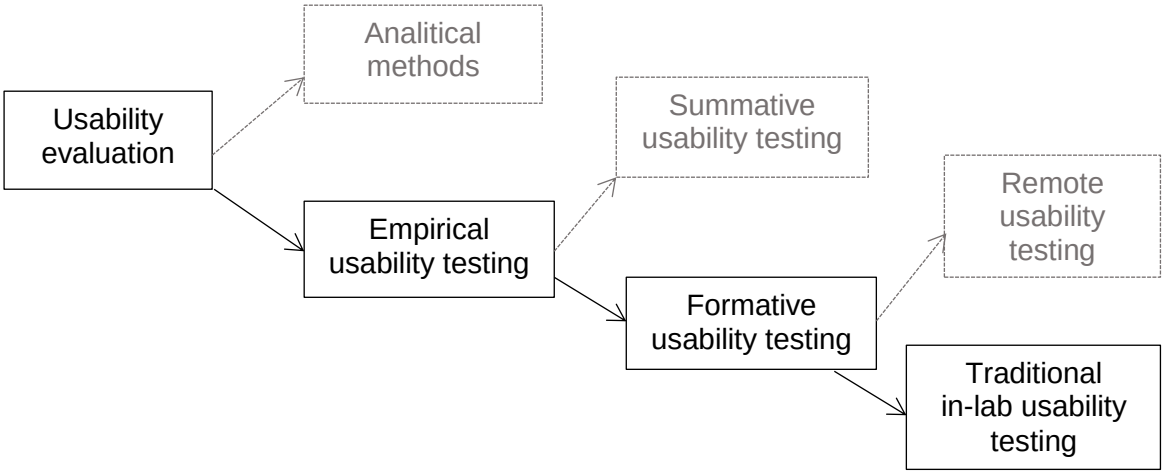


Figure 8. Choosing an appropriate approach for usability evaluation of Lioncrypt 1.0

After the careful consideration, the complementary combination of methods and techniques was proposed to exploit the full potential of usability testing approach. The scenario-based design of usability testing was chosen to replicate the natural environment of software usage. The opportunity to observe the potential users performing realistic tasks were found to be of great significance.

Further, such techniques as concurrent thinking aloud and *observation* were adopted to collect the qualitative data. Both procedures (see subchapters 4.3 and 4.4 for a complete discussion*)* enable to collected thoughts, emotions and expressions from users, hence, help to understand the way of thinking of the users.

Concurrent thinking aloud (CTA) was chosen over retrospective think aloud (RTA), because of several reasons. In the first place, each test session was supposed to last a relatively long period of time, and distinct advantage of CTA is that it is performed simultaneously while fulfilling the tasks. To repeat, RTA can significantly prolong the test session time, which might lead to frustration of the participants. Another reason is that during CTA participants verbalize their thoughts immediately, in contrast, while using RTA they could forget specific details.

Data *logging* during usability testing*,* used with observation and thinking aloud protocol, allowed to document participants behavior and vocalized thoughts. As mentioned in subchapter 4.4, logging can be done manually or using special software. In the current

study, in pursuit of staying within the boundaries of the budget of the firm, data logging was deployed using Microsoft Excel observation form with time stamping feature. Additionally, logging was complimented with basic note taking.

Speaking of questionnaires, the well-known, free and reliable *SUS* and *SEQ* were utilized in this study. As was previously stated (see subchapter 4.5), these particular questionnaires are universal for any usability research, regardless of its size.

As previously discussed in section 4.5.3, the main goal of SEQ is to capture the participants feeling immediately after the completion of each task. Also, the idea behind including SEQ after each task is to help participants to cope with boredom from carrying out a long set of tasks. The SUS questionnaire was chosen for the post-test usability evaluation with two minor modifications:

- the word “cumbersome” was replaced with the word “awkward” in statement 8
- an open-ended question was added at the very end of the SUS questionnaire.

The first-mentioned adjustment to the SUS questionnaire was applied in order to avoid words that can create confusion for non-native English speakers. The open-end question: “What would you improve?” was attached to SUS questionnaire to investigate personal opinions that might have remained unexposed during the usability testing.

Another argument for the choice of methods and techniques discussed above is that they are applicable for small studies and allow to collect *quantitative* and *qualitative* data (see chapter 4). Figure 9 illustrates the final selection of methods and techniques, related to the traditional in-lab usability testing, chosen for the usability evaluation of Lioncrypt 1.0.

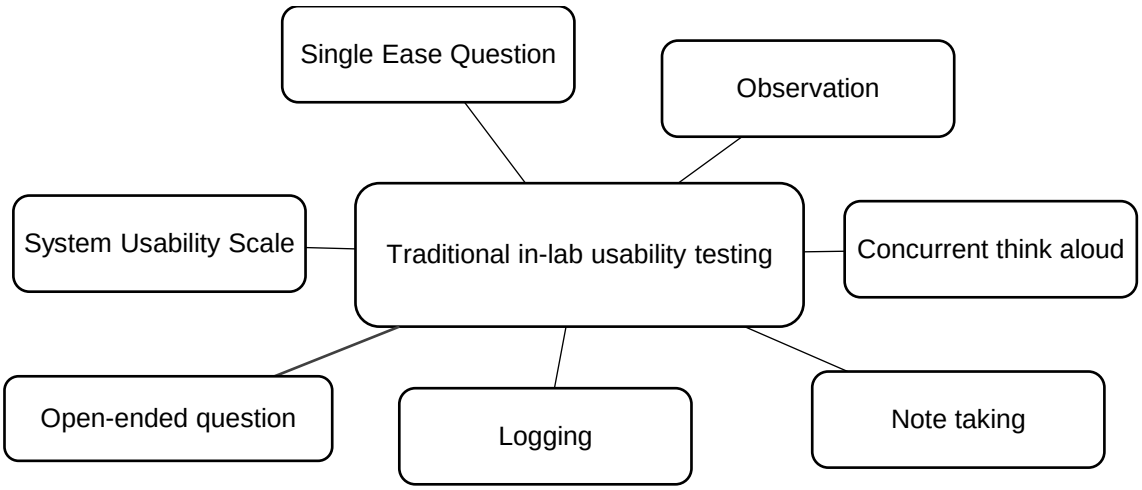


Figure 9. Methods and techniques proposed for this study

5.3 Participants

As discussed in subchapter 4.6, for formative usability testing recruiting as much as 5 participants, as long as they represent one subgroup of the target audience, generates valid results.

Based on *persona CEO* (appendix 1), created before the study was undertaken, characteristics for selection of the participants have been determined as follows:

- Familiarity with the product: novice Lioncrypt user – used Lioncrypt from 0 to 5 times.
- Computer skills: competent/proficient user.
- Computer usage at least 20 hours per week.
- Education: BBA, MBA or higher.
- Job category: CEO, director, president.
- Company size: micro to small company.
- He has concerns about data security.
- Dropbox usage.

The candidates with an advanced level of English and renting the office spaces in same the business center as the company under consideration were encouraged to take a part in usability testing of a Lioncrypt 1.0 software. Potential participants were contacted directly through Skype, email, and verbal communication.

Volunteered participants were invited to fill in the *Participant screening questionnaire* (appendix 2) online through Google Forms (URL: http://goo.gl/forms/r9N9U5eVb5) to fit the characteristics of the chosen group of the representative users (see section 4.5.1). Five participants, matching the defined criteria, were recruited before testing.

5.4 Environment and equipment

The usability evaluation of Lioncrypt 1.0 was performed in the conference room, set up as an informal lab. The environment for the controlled traditional usability testing is illustrated in figure 10.

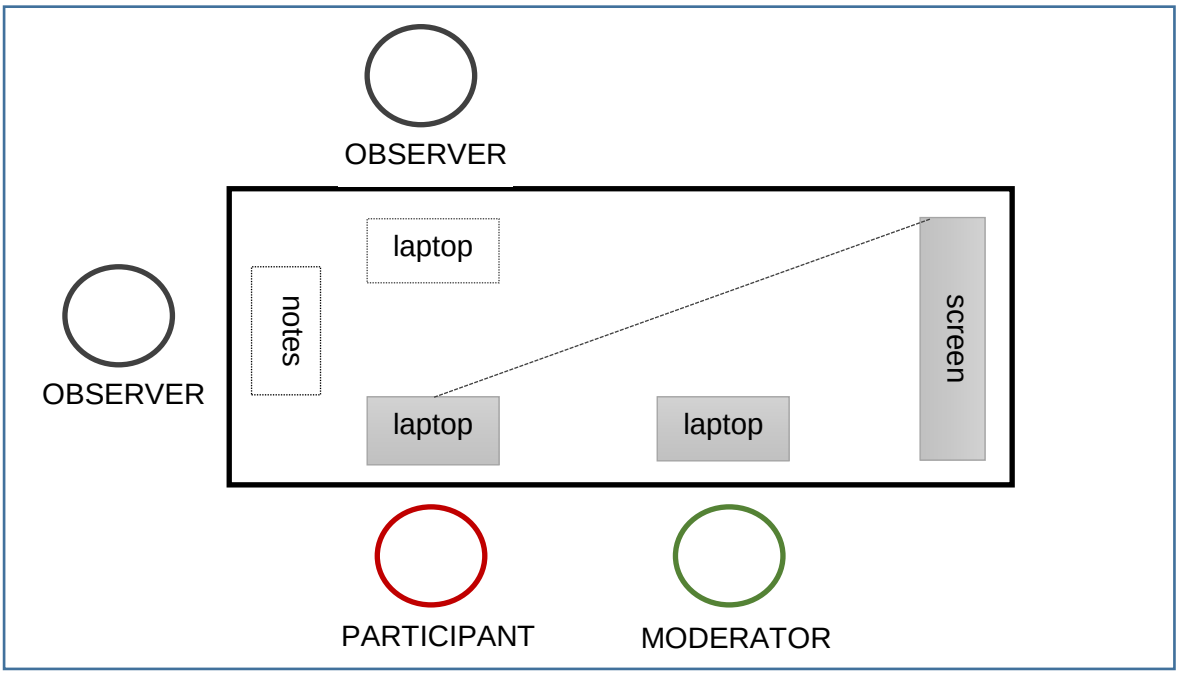


Figure 10. Usability testing environment and observation set-up

Non-technical equipment:

- desk
- chairs
- note paper and pen.

Technical equipment and software:

- laptop for the participant with a microphone for audio recording
- laptop for the moderator
- large display, connected to the laptop of the participant (to project screen activity).

In addition, free Open Broadcaster Software (OBS) was utilized for the purpose of recording audio and screen actions performed by the participant. Moreover, data logging was handled using Microsoft Excel spreadsheet.

5.5 Procedure

The moderated in-lab usability testing was carried out during the period from 25.03.2015 to 02.04.2015 in the conference room of Ferus Bestia Oy in Lautasaari (Helsinki). The study comprised five test sessions, individual for each participant. Also, the main study was preceded by a pilot test, aimed to ensure that test processes and materials are ready for the usability evaluation study. However, one of the participants could not take a part in the research. Still, the data received from the pilot session participant is considered reliable due to the reasons that the participant represented the target audience, and the

scenario has not been changed (Barnum 2011, 251). The duration of test session ranged from one to two hours, depending on the time required for a participant to complete all tasks. The valid testing schedule can be found in appendix 3.

Each test session was guided by the *moderator*, who read out loud the task description at the start of each task, interacted with the participant and remind him to keep talking aloud. Additionally, the moderator played the role of the data logger. Data logging observation form and logging codes are presented in appendix 5 and appendix 6. Aside from the moderator, at least one *observer* (the development team members, CEO of the company) was invited to watch the test session and take notes. Also, a *technician* provided technical support during the test.

Upon arrival at the scheduled test session, each participant was offered to sign a consent and recording release form (appendix 4), and proceed with the familiar environment: Mac (Osx) or PC (Windows). Besides, the participant was allowed to use the browser of choice. Then, the participant was briefly instructed about test objectives and processes, think-aloud protocol and Lioncrypt 1.0. After that, the moderator provided the participant with the main scenario (appendix 8), elaborated in regard the predefined set of tasks (appendix 7). The scenario was divided into 8 separate scenarios + 1 extra scenario, due to the reason that developers had doubts about including the extra scenario. Nevertheless, it was successfully used during each of the performed test sessions. Each participant used the same scenario to keep consistency. Scenario completion rate of 100% is the goal for each scenario.

Along with thinking aloud, the participant was required to indicate start and end points of each task for more accurate data collection. Upon completion of each task, the participant was asked to fill post-task questionnaire (SEQ), provided within the scenario. Eventually, at the end of all the scenarios, the participant was asked to fill in the post-test SUS questionnaire with the addition of open-end question (appendix 4).

6 Results and Discussion

This chapter presents the outcome of the usability evaluation study, which was described in the previous chapter. This chapter is organized as follows: the first section illustrates the data collected about the recruited participants, the order of the other results presentation will follow the research questions mentioned before.

6.1 Participants profile

For the usability evaluation of Lioncrypt 1.0 as much as five participants were recruited for the usability testing procedure. Two out of three of participants were aged between 20 and 30 years (see figure 11). Also, the majority of participants held an undergraduate or bachelor degree (see figure 12).

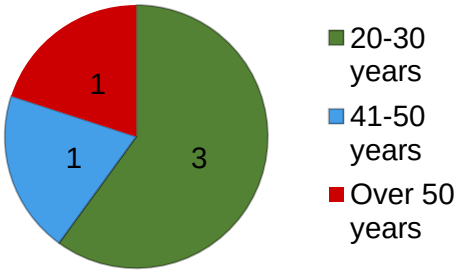


Figure 11. Participants by age (N=5)

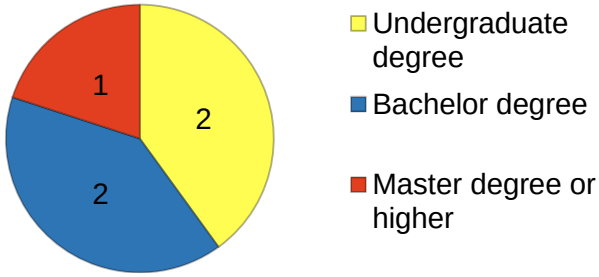


Figure 12. Participants by educational background (N=5)

All participants outlined their occupation as CEO/Managing director of the company, which has fewer than ten employees. Also, all of the respondents were male and non-native English speakers. Familiarity with Dropbox and an intention to improve security was also in common for all the participants.

Two out of five participants had experience with Lioncrypt software, however it has not exceeded five times. Full participant summary data can be found in appendix 9.

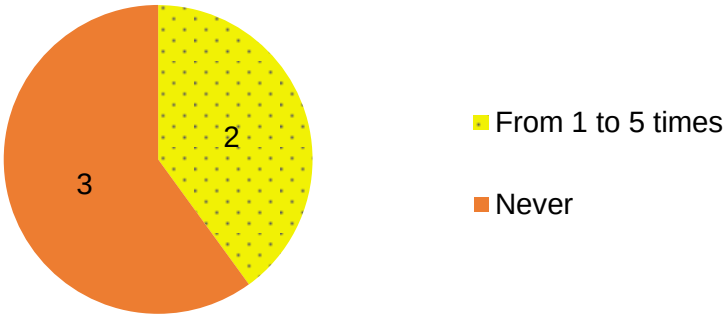


Figure 13. Familiarity of the participants with Lioncrypt 1.0 (N=5)

6.2 Attitude of the participants towards the use of Lioncrypt 1.0

The overall impression of the participants from using Lioncrypt 1.0 is the first priority criteria. The quantitative data, derived from the SUS questionnaire was analyzed in order to investigate, how easy and pleasant was the interaction with Lioncrypt 1.0 for the participants. In attempt to explore this issue more fully, traditional and two-dimensional approach to the SUS score calculation were employed (see subchapter 4.5.2 for more details). The following figure 14 illustrates *traditional approach* to SUS score results calculation.

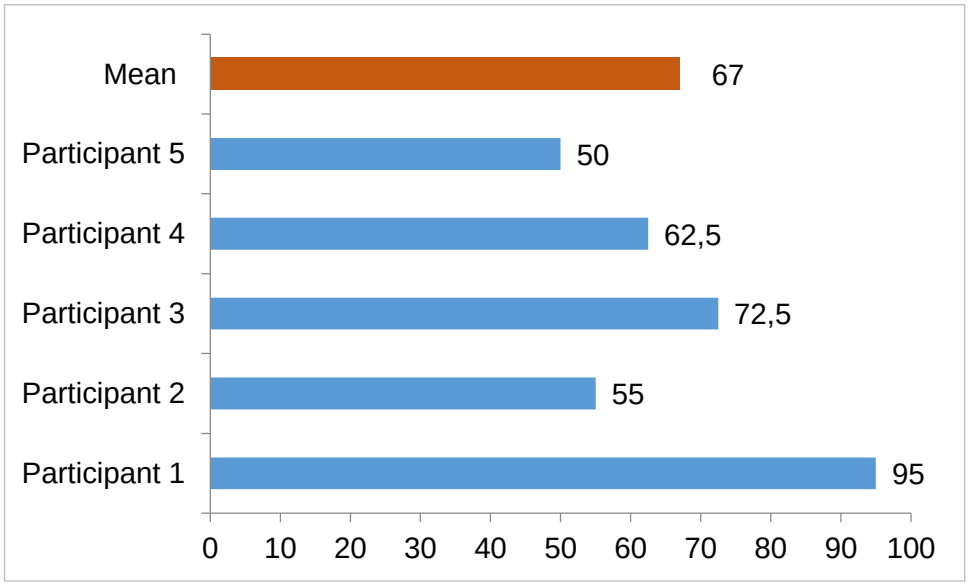


Figure 144. SUS scores by participant (N=5)

As can be seen from the graph, participants SUS scores noticeably vary. Interesting fact that the older respondents rated overall impression of the usability of Lioncrypt 1.0 lower, then participants from the age group of 20 to 30. This finding needs to be tested in future research, most likely with extended sample size.

Also, the SUS contribution scores of participant 1 and participant 3 are distinctly higher. That means, they found the overall usability UI of Lioncrypt 1.0 higher, than other respondents. This finding probably correlates with the fact that participant 1 and participant 3 previously had an experience of using the software. Which supports the pattern discovered by Sauro (2013d) that the most experienced users rate websites and software with higher SUS scores.

Summing up, the mean SUS score, calculated using traditional approach is 67, which is a little below average (Sauro 2011b). Also, according to above mentioned Bangor & al. (2009, 119), the mean SUS score belongs to “ok” or “so-so” rating interval. With regard to

two-dimensional approach to SUS score calculation, the usability mean SUS score (calculated with relation to learnability) is 64, 38, which is similar to the previous result. And again, mean SUS score belongs to “ok” rating interval. All in all, the result above enable to draw a conclusion that the usability of Lioncrypt 1.0 is not good enough. However, it cannot be labeled as software with poor usability.

6.3 Ease of learning for novice users

In order to assess how easy is learn how to use Lioncrypt 1.0 for novice users, *two-dimensional approach* (see subchapter 4.5.2) of calculating mean SUS score was utilized. The learnable mean SUS score of Lioncrypt 1.0 has reached 77,5 on a scale from 0 to 100. Interpreting this value using Bangor & al. (2009, 119) rating scale, the learnability of Lioncrypt 1.0 can be estimated as “good”. In practical terms, this means that the novice participants could quite easy learn how to use Lioncrypt 1.0.

6.4 The most difficult scenarios for novice users to accomplish

Aiming to investigate, which scenarios were the most problematic to complete, both performance (based of quantification of actions) and satisfaction (based on preferences) measures were analysed. Performance measures are represented by time-on-scenario and scenario success rate. Whereas results derived from SEQ questionnaire serve as satisfaction measure.

6.4.1 Time-on-scenario

Time-on-scenario quantitative data was derived from the log files, which encompassed observation and thinking aloud data. The timestamp from each observation entry was converted into seconds.

During the current study, one of the participants accomplished each scenario much longer than the others. This phenomenon was described by Barnum (2011, 254) as an obstacle for presenting mean time-on-task and summarizing quantitative findings in small studies. However, it is recommended to keep the results received from testing the so-called “wildcard” on time participant as he could represent 20% of target users in the study of 5 participants. Also, it is advisable to present the time-on-task results as a table, containing the actual times per each participant. (Barnum 2011, 254.)

Table 1. Time-on-scenario data per each participant in seconds (N=5)

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Scenario 1	378	2263	214	318	284
Scenario 2	122	886	197	349	299
Scenario 3	170	1056	215	337	575
Scenario 4	302	2191	227	265	900
Scenario 5	88	328	34	79	29
Scenario 6	107	288	43	115	67
Scenario 7	44	251	26	49	48
Scenario 8	111	122	70	312	392
Extra Scenario	132	751	162	180	491

It can be clearly seen that for all the participants (including the “wildcard”) Scenario 1 and Scenario 4 appeared to be the most challenging, as they spent the most amount of time to complete them.

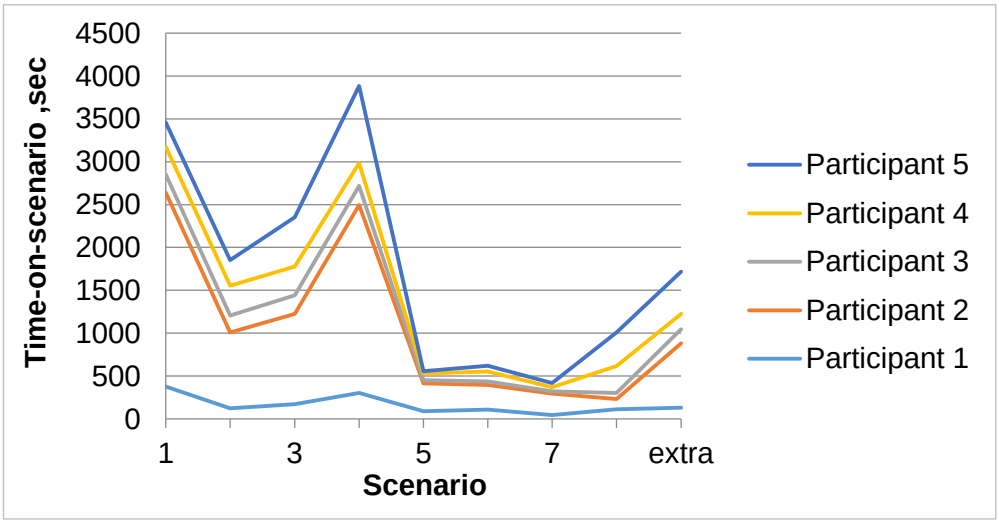


Figure 155. Time-on-scenario per each participant (N=5)

6.4.2 Scenario success rate

These quantitative data, concerning the success of accomplishing the scenarios by the participants, were retrieved from the log files. The results summary is provided in appendix 10. The results were analyzed in terms of *success*, *partial success* and *failure*. *Success* refers to tasks, accomplished by the participant without any involvement of the moderator (verbal or non-verbal). Success is rated as 100%. The term *partial success* means that the moderator read the scenario to the participant once again or gave a slight hint about the process. Partial credit of 50% is granted to partial success. Finally, failure refers to the tasks, assisted by the moderator and is rated as 0%. This approach was borrowed from Nielsen (2001).

Table 2. Success completion rate for 5 participants (P – participants, S – Scenario)

	S1	S2	S3	S4	S5	S6	S7	S8	Extra S	Mean
P1	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
P2	50 %	0 %	0 %	0 %	50 %	50 %	50 %	100 %	0 %	33 %
P3	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
P4	100 %	100 %	100 %	50 %	50 %	100 %	100 %	100 %	100 %	89 %
P5	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	50 %	94 %
Mean	90 %	80 %	80 %	70 %	80 %	90 %	90 %	100 %	70 %	83 %

The above table exhibits that the “wildcard” participant 2 showed the worst level of scenario completion. In contrast, the majority of participants completed almost all the scenarios successfully. Actually, average scenario completion rate of the participant 2 was only 33%, while other participants reached the level from 89% to 100%. Mean scenario completion rate per all participants is 83%. Even though the result of the “wildcard” stands out from the others, it had a minor impact on the success rate of each scenario as demonstrated in the graph below.

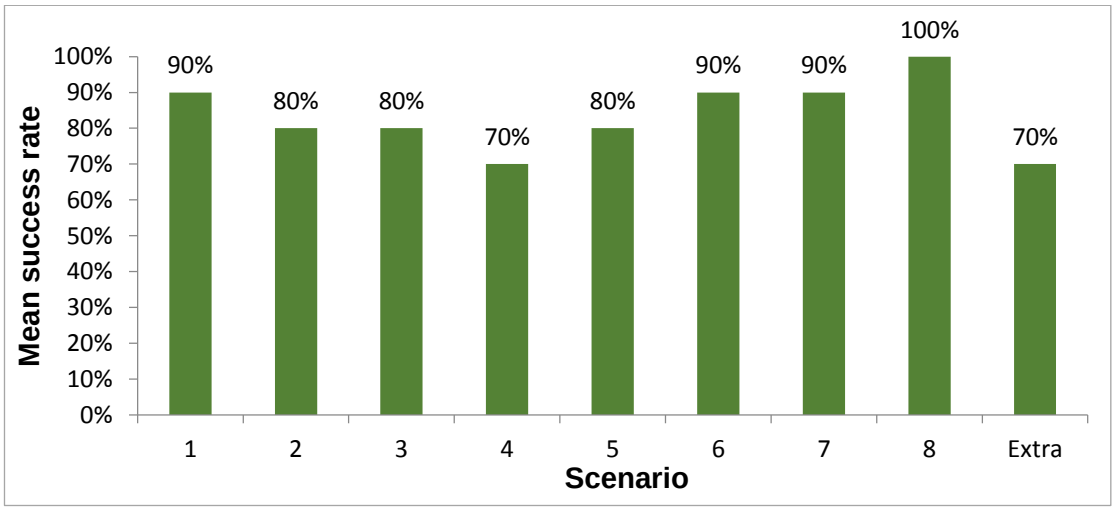


Figure 166. Success completion rate per each scenario (N=5)

As can be seen from the graph, the most difficult scenarios for participants to complete were Scenario 4 and Extra Scenario. Also, for Scenario 2, Scenario 3 and Scenario 5 the mean success rate is slightly lower, then for the rest.

6.4.3 SEQ results

In contrast with considered in previous two sections performance measures, responses on the SEQ refer to post-task satisfaction. This quantitative feedback was collected using the 7-point scale, rated by participants after each scenario. Due to the reason that one data set was lost due to the human error, the mean SEQ score was calculated for 4 participants.

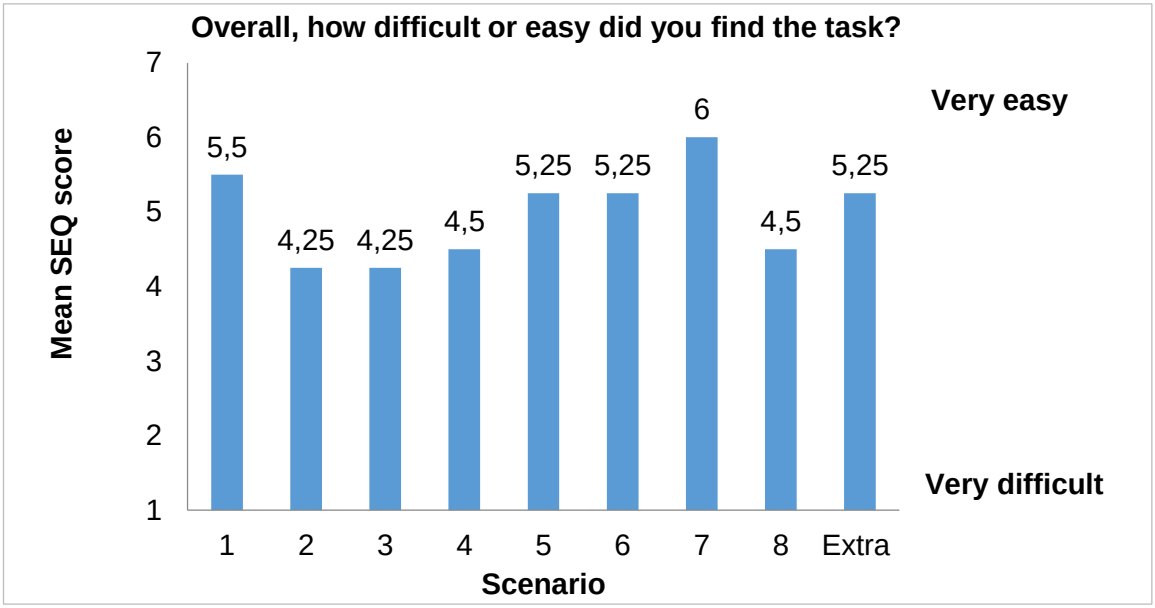


Figure 177. Mean SEQ score (N=4)

From the above graphs, it can be seen that participants did not feel that the tasks were difficult. The lowest scores were given to Scenario 2 and Scenario 3. Also, Scenario 4 and Scenario 8 scores are a little lower than the industry average (see subchapter 4.5.3). That means, the participants found Scenario 2 and Scenario 3 the most difficult. Also, participants were of the similar opinion about Scenario 4 and Scenario 8.

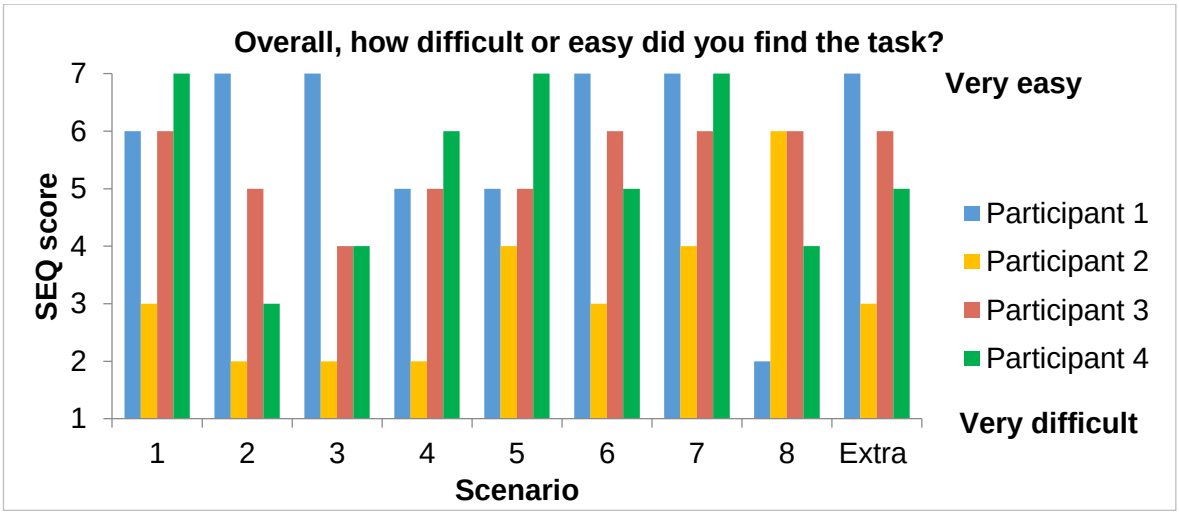


Figure 188. SEQ score per each participant (N=4)

As mentioned earlier in this chapter, each participant has a high value in small studies. Therefore, to add validity and reliability to the results, SEQ scores were also presented for each participant. From the Figure 18 we can see, that the impressions about difficulty of tasks collected from so-called “wildcard” Participant 2 reflected his performance difficulties. While other participants mostly ranged tasks similar to average industry benchmark.

6.4.4 Difficult scenarios matrix

To sum up the above results, the matrix of difficulties has been developed based on performance measures and user self-assessment.

Scenario	1	2	3	4	5	6	7	8	Extra
Time-on-scenario									
Success rate									
SEQ									

Figure 19. The matrix of difficulties in accomplishing scenarios (N=5)

Clearly, Scenario 4 was found the most difficult for novice users to accomplish. In this particular scenario, each participant was asked to create Lioncrypt folder, encrypt files and share this folder with another Lioncrypt user. This might be explained by the fact that participants, perhaps, misunderstood the idea behind Lioncrypt folders concept and Lioncrypt sharing feature (see subchapter 6.5 for more details about these problems). This finding might require further research.

Scenario 2 and Scenario 3 also demand special attention. During Scenario 2 each participant was asked to connect his Dropbox account to his Lioncrypt account. Most likely, the participants experienced difficulties, because they feel disoriented due to flaws in the website UI navigation and the colours. This issue is also discussed in the subchapter 6.5.

Finally, in Scenario 3 participants had to download and install the appropriate version of Lioncrypt application. Difficulties in completing this scenario were probably caused by the fact that, at first, participants had to log out and then download the Lioncrypt desktop application from the website.

6.5 Usability problems and possible solutions

In this subchapter qualitative data, which was collected through the think-aloud protocol, data logging, note taking and open-end question at the end of SUS (“What would you improve?”) is presented and discussed.

The development team decision was to avoid providing the training for the participants before the testing, except some general information about Lioncrypt 1.0. The idea behind this solution was to get a glimpse into the way of thinking of the respondents without any experience or with minimal experience while they try to achieve their goals.

During the study, negative and positive quotes, questions and improvement suggestions for each step of the scenario were collected. Based on this comments, the most common and important problems were identified and presented below.

Also, in order to weight the usability problems, the rating of severity, proposed by Sauro (2013b), was adopted:

- *Minor* problem acts as a source of hesitation or slight irritation.
- *Moderate* leads to irritation, delay and scenario failure of some participants.
- *Critical* rating indicates that the problem leads to task failure and cause a high level of irritation.

Table 3. Problems with the number of affected users and severity rating

N	UI	Number of affected participants	Scenario	Problem for participants	Severity
1	Website	3	Scenario 1	Sharing option during adding new users	Minor
2	Website	2	Scenario 1	Expiring option during adding new users	Minor
3	Website	3	Scenario 2	Difficult to find how to add Dropbox	Moderate
4	Desktop application	3	Scenario 3	Difficulty to find desktop application after the installation	Moderate
5	Desktop application	3	Scenario 4 Scenario 5	Annoying pop-ups and small letters	Minor
6	Desktop application	1	Scenario 6	Cannot make spaces in Lioncrypt folder names	Minor
7	Desktop application	1	Scenario 6	Lioncrypt folder disappeared after creation	Critical
8	Desktop application	2	Extra scenario	Typo in the confirmation question	Minor
9	Website	4	Scenario 1 Scenario 2 Scenario 4	Confusion about admin-users concept	Moderate
10	Desktop application	3	Scenario 4-8 Extra Scenario	Confusion about Lioncrypt folders concept	Moderate
11	Desktop application	3	Scenario 4-8 Extra Scenario	Confusion about sharing feature	Moderate
12	Website and desktop application	4	All	Misunderstanding of the concept website-desktop application	Critical

The following graph, representing the frequency of participants affected by specific usability problems was generated based on table 3.

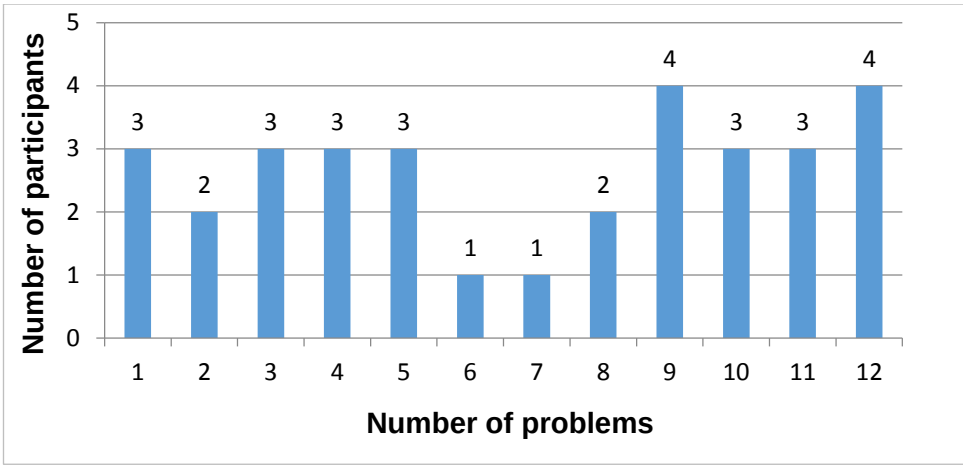


Figure 20.Frequency of participants affected by one of 12 uncovered usability problems (N=5)

The uncovered problems are discussed below in more detail. A screenshot is provided when necessary in order to portray the problem. Also, the *digits* are used to display problem frequency for illustrative purposes.

***Problem 1. Sharing option during adding new users*.**

3 out of 5 participants were confused about sharing option during adding new users while using website UI (see Figure 21). An example of typical comment would be the following: “Share folders, what does it mean?” The meaning of this option was that the Admin user could provide the rights to share folders to the normal users of the Lioncrypt 1.0. For example, for the users who do not continue their contract, the option to share Lioncrypt folders would be unavailable.

Recommended fix: for the Lioncrypt 1.0 this option might be a priori set as positive, as for MVP version the overall financial strategy was not fully established. In this case, this multi-choice alternative would be removed from the website UI.

Problem 2. Expiring option during adding new users.

2 out of 5 participants asked, what are the consequences if the account expire. In addition, 1 of the respondents suggested using hints, which would explain the meaning of this option. Figure 21 demonstrates the screenshot overview of the first two problems. *Recommended fix:* some short explanation can be added to website UI to explain the meaning. Other solution would be to remove this option as insignificant.

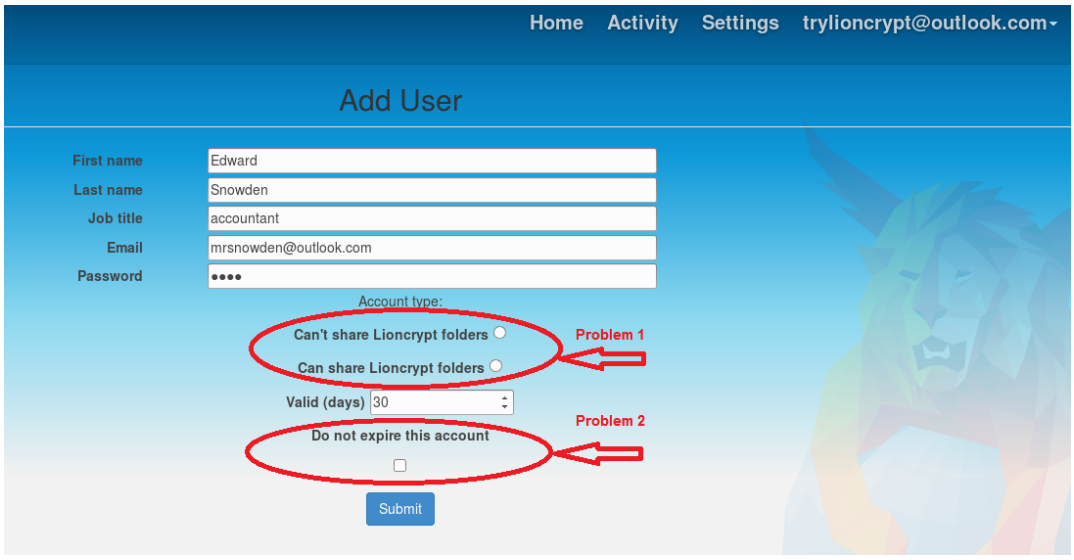


Figure 21. Adding a new user (Lioncrypt 1.0 website UI)

Problem 3. Difficult to find how to add Dropbox.

3 out of 5 participants indicated a frustration from the attempt to link the Lioncrypt account of a user to the Dropbox. Notable comments:

- "Not easy to find."
- "Button color of the Dropbox is not visible."
- "I cannot see where to connect."

As can be seen from the figure 22, the background of the website UI and the button "Add Dropbox" are of different variations of the color blue. Hence, the button does not stand out of the background. Additionally, participants normally undertook several attempts to figure out where is this option placed.

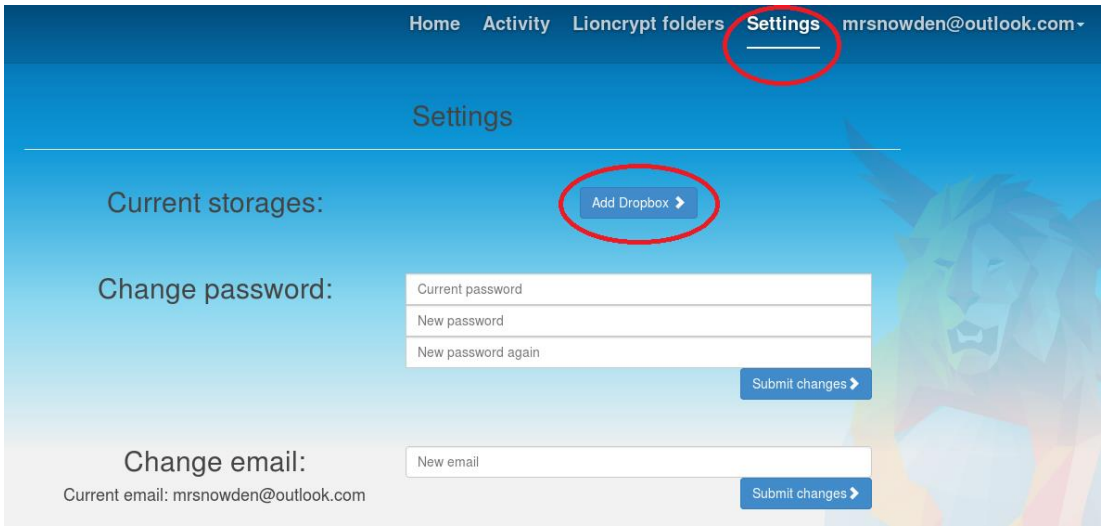


Figure 22. Linking Dropbox account to Lioncrypt account (Lioncrypt 1.0 website UI)

Recommended fix: change the button color to emphasize the contrast. Move linkage to Dropbox account from underneath Settings to the first screen, which appears after a user logs in.

Problem 4. Difficulty to find Desktop application after the installation.

3 out of 5 participants reported difficulty in finding Lioncrypt 1.0 desktop application after its installation. Comments related to the issue:

- “Not visible on screen.”
- “Where is it.”
- “This should have a logo.”

Recommended fix: add a logo to Lioncrypt 1.0 desktop icon.

Problem 5. Annoying pop-ups and small letters.

3 out of 5 participants were irritated by the pop-up progress bar. 2 of them mentioned, that the letters are probably too small. Figure 23 shows a progress bar during logging to Lioncrypt 1.0 desktop application.

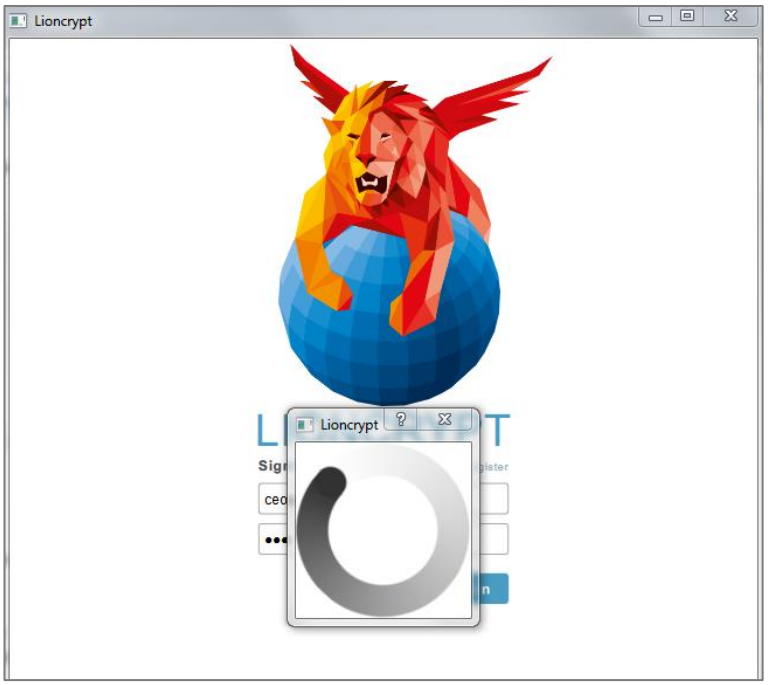


Figure 23. Logging to Lioncrypt 1.0 (desktop application UI)

Comments, collected through think-aloud, logging and notes taking:

- “Looks weird, this up and down logging.”
- “Funny and annoying.”
- “Letters are too small.”

Recommended fix: change current progress indicator to another one.

Problem 6. Cannot make spaces in Lioncrypt folder names.

1 out of 5 participants noticed that it is not possible to make spaces in the name of Lioncrypt folder.

Recommended fix: might be considered as an option to add.

Problem 7. Lioncrypt folder disappeared after creation.

1 out of 5 participants reported that if you do not set the password for a Lioncrypt folder during its creation it disappears afterward. That means the Lioncrypt folder was not actually created in this case. This problem does not really fall into the category of usability problems, most likely it should be reported as a bug.

Recommended fix: find a technical solution to the bug. This is a high priority issue and should be fixed by the development team as soon as possible.

Problem 8. Typo in the confirmation question.

2 out of 5 participants noticed that the confirmation message in extra scenario has a typo. A comment example: “The word “accept” is with three c”.

Recommended fix: correct the typo.

Problem 9. Confusion about admin-users concept.

4 out of 5 participants misunderstood that Admin account is used only to perform administrative functions. Even though the brief explanation was provided in the beginning of each testing session, participants instantly forgot the instructions.

Recommended fix: explain this concept in detail or provide written/video instructions.

Problem 10. Confusion about Lioncrypt folders concept.

3 out of 5 participants did not understand, why they have to create Lioncrypt folder in order to add files for encryption. This concept is the core component of the software, and it most likely cannot be changed.

Recommended fix: explain this concept in detail or provide written/video instructions.

Problem 11. Confusion about sharing feature.

3 out of 5 participants experienced difficulties with the understanding of the sharing feature. Meaningful comments:

- “How can I see if the folder is shared?”
- “Why it does not say that it is shared?”
- “It is misleading to not tell that it is shared”.

Recommended fix: add necessary notifications and status.

Problem 12. Misunderstanding of the concept website-desktop application.
4 out of 5 participants were disoriented when working with 2 different interfaces.
Recommended fix: provide only single UI for interaction, probably desktop application.

6.6 Positive findings and suggestions

Presenting only negative study results might upset and discourage the members of the development team. Therefore, it is recommended to present positive findings along with the negative ones. (Sauro 2013b.) In this research positive findings and suggestions were mostly collected through think aloud protocol and SEQ. This data was documented via note taking and logging, together with audio recording.

To begin with, some positive finding occurred during the analysis provided in the previous chapters. For example, Scenario 8 was successfully completed by all participants without assistance. This scenario was asking to share files with another Lioncrypt user. Given the fact that the similar Scenario 4 was found the most difficult to complete, it might be assumed that the participants were able to learn how to interact with the software.

Participants expressed positive opinions about the colors of website UI of the Lioncrypt 1.0 and Lioncrypt lion logo. Scenario 4, where participants added files to the Lioncrypt folder, was found the easiest of all. Participant 3 commented that files were added to Lioncrypt folder very fast (Scenario 4). Participant 4 was impressed by sharing feature with another Lioncrypt user (Extra Scenario).

Speaking of suggestions, one participant mentioned that he would like the design of Lioncrypt UI to be similar to Google Drive. Other suggestions are summarized below:

- Change the color of Dropbox button to green (Lioncrypt 1.0 website UI).
- Add a logo to the Lioncrypt 1.0 desktop application.
- Having less different passwords (website and desktop Lioncrypt 1.0 UIs).
- Changes to the progress bar in Lioncrypt 1.0 desktop application (for example, “there could be some text about encrypting or decrypting files”).
- The sharing status should be provided in the Lioncrypt 1.0 desktop application.

6.7 Reliability and validity of the results

To ensure the reliability of qualitative results, the data was obtained through observation, think-aloud protocol and open-ended question. In addition, the data was collected by different people, for example, moderator and observer. The analyses of collected data

demonstrated the consistency of results, hence, the reliability. The usage of reliable tools, such as SUS questionnaire and SEQ (Sauro 2010; Sauro 2011b), known to produce valid measures, facilitated to obtain reliable qualitative results.

For present research one limitation would be a small sample size. However, for formative usability testing a sample size of five participants, selected from one subset of the target audience is generally enough to obtain valid results (Wiklund & al. 2016, 115), this topic is thoroughly discussed in chapter 4. The screening questionnaire used in this study attempted to select participants from only one group of potential users of Lioncrypt 1.0. Although the participants most likely were the representatives of the target users group, due to the size of the sample, this results might not be reliable with regard to a larger group of the target audience (Hughes 2011). Therefore, this results should be analyzed in conjunction with other results, obtained during the iterative testing process.

As stated by Walliman (2011, 45): “If you can see any possibility of bias in any aspect of the research, it should be acknowledged and explained.” In present research bias might arise, for instance, due to the data loss from SEQ answered by one of the participants. However, the present formative usability testing study is primarily focused on qualitative data, and the results from this very first testing are not statistically significant. Therefore, this result should be considered in the context and in conjunction with the results from the following evaluations.

Another thing to consider is the scenario design and description. As mentioned earlier, scenario-based usability testing in the informal lab was chosen to imitate natural environment of Lioncypt 1.0 usage. Nevertheless, unnatural factors, such as thinking aloud (CTA) and the presence of moderator/observer, have an influence on the behavior of the participants. The artificial environment could lead to bias in responses and reactions Walliman (2011, 104). Opinions about so-called “observer effect” contradict each other so greatly, that this topic was not covered in this thesis. That might be a subject of further research.

Finally, the accuracy of scenarios might have an impact on validity. The clarity of directions should be further investigated. However, the fact that the majority of participants did ask questions about the listed goals can be seen as a positive sign.

7 Conclusion and future recommendations

In recent years, the beneficial impact of integrating the usability evaluation into software development cycle has been widely acknowledged. However, many companies delay or omit this procedure due to a lack of human and financial resources. Ignoring the usability issues might undermine the software success on the market. Obviously, it was one of the main concerns of the Ferus Bestia Oy, which was prepared to release a new software - Lioncrypt 1.0.

The multi-method framework constructed for usability evaluation of Lioncrypt 1.0 enabled to answer to the main concerns of the development team. Empirical usability testing with five novice users of Lioncrypt 1.0 allowed to receive feedback directly from the potential users of the software and stay on the budget via setting up the conference room as an informal lab.

The think-aloud protocol and observation techniques, supported by logging and note taking, appeared to be a fruitful source of qualitative data. The answers to SEQ and time-on-scenario represented numeric data. The evaluation study findings revealed a number of difficulties and 12 usability problems related to desktop and website user interfaces. One of the problems most likely can be considered as a bug. Almost a half of the problems reached quite high severity rating. That means they have to be resolved before the release of Loncrypt 1.0. The list of possible solutions was offered to cope with these issues.

Several interesting findings emerged from the analysis of quantitative data, derived from the tailored post-test SUS questionnaire. Overall usability of Lioncrypt 1.0 software was estimated as “slightly below average”. In the same time, the level of learnability, according to the SUS results was above “good” level. Even though the participants did not find Lioncrypt 1.0 very easy and pleasant to use, they pointed out that it was easy to learn how to use the software.

In the light of above, it can be clearly seen that the developed framework has proved to be useful in assessing the usability of Lioncrypt 1.0 and identifying usability problems. Integrating this framework into iterative development process would constitute a valuable asset for Ferus Bestia Oy. Also, due to its effectiveness and cost-effective manner, this framework might be adapted by other researchers and serve as a basis for usability evaluation of various software.

Future recommendations would include the elimination of the discovered usability violations, followed by repeating the usability evaluation. Then, results of these studies can be compared in order to analyze an impact of the applied improvements. It is highly recommended to collect both qualitative and quantitative types of data as they can positively complement each other. In addition, it is advisable to analyze the quantitative data collected through several study iterations. On this condition, confidence interval should be calculated in order to estimate the accuracy of quantitative results. Also, repeating the same study design can increase the reliability of the framework. Another factor for the further research might be “observer effect”, as it might influence the reliability of the results.

It is important to remember that the purpose of usability evaluation is to improve the communication between user and software system. My belief is that there is always room for improvement. Therefore, usability evaluation should be considered as an iterative and never-ending process.

References

Albert, B. & Tullis, T. 2013. Measuring the user experience: collecting, analyzing and presenting usability metrics. 2d ed. Morgan Kaufmann Publishers Inc. San Francisco.

Albert, B., Tullis, T. & Tedesco, D. 2010. Beyond the usability lab: conducting large-scale online user experience studies. Morgan Kaufmann Publishers Inc. Burlington.

Andreasen, M., Nielsen, H., Schroder, S. & Stage, J. 2007. What happened to remote usability testing? An empirical study of three methods. In CHI '07: proceedings of the SIGCHI conference on human factors in computing systems, pp. 1405-1414.

Ardito, C., Costabile, M., De Angeli, A. & Lanzilotti, R. 2006. Systematic evaluation of e-learning systems: an experimental validation. In NordiCHI '06: proceedings of the 4th Nordic conference on human-computer interaction: changing roles, pp. 195-202.

Baguma, R, Kiprono, K. & Kirui, E. 2016. Usability framework for E-Government services in developing countries. International Journal of Computer Applications, 137, 11, pp. 43-47.

Baker, R. 2014. Remote usability testing: thinking outside the lab. UXmatters. URL: http://www.uxmatters.com/mt/archives/2014/04/remote-usability-testing-thinking-outside-the-lab.php. Accessed: 31 March 2016.

Bangor, A., Kortum, P. & Miller, J. 2009. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability studies, 4, 3, pp. 114-123. URL: http://uxpajournal.org/wp-content/uploads/pdf/JUS_Bangor_May2009.pdf. Accessed: 14 March 2016.

Barnard, Y., Sandberg, J. & Someren, M. 1994. The think aloud method. A practical guide to modelling cognitive processes. Academic Press. London. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.7738&rep=rep1&type=pdf. Accessed: 20 April 2015.

Barnum, M. 2011. Usability testing essentials: ready, set...test! Elsevier Inc. Burlington.

Battleson, B., Booth, A., & Weintrop, J. 2001. Usability testing of an academic library Web site: a case study. Journal of Academic Librarianship, 27, 3, pp. 188-198.

Bergstrom, J. & Olmsted-Hawala, E. 2012. Think-aloud protocols: does age make a difference? Academia.edu. URL: http://www.academia.edu/1241322/Think-Aloud_Protocols_Does_Age_Make_a_Difference. Accessed: 16 March 2016.

Bevan, N. 2008. Classifying and selecting UX and usability measures. COST294-MAUSE Workshop: meaningful measures: valid useful user experience measurement. Reykjavik. Nigel Bevan, professional usability servicies.URL: http://www.nigelbevan.com/papers/Classifying%20and%20selecting%20UX%20and%20usability%20measures.pdf. Accessed: 11 March 2015.

Bevan, N. 2009a. International standards for usability should be more widely used. Journal of Usability studies, 4, 3, pp. 106-113. URL: http://uxpajournal.org/wp-content/uploads/pdf/JUS_Bevan_May2009.pdf. Accessed: 09 April 2016.

Bevan, N. 2009b. What is the difference between the purpose of usability and user experience evaluation methods. UXEM'09 Workshop. INTERACT 2009. Uppsala.URL: http://www.nigelbevan.com/papers/What_is_the_difference_between_usability_and_user_experience_evaluation_methods.pdf. Accessed: 11 March 2015.

Bevan, N., Carter, J. & Harker, S. 2015. ISO 9241-11 revised: What have we learnt about usability since 1998? In Kurosu, M. Human-computer Interaction, 1, pp. 143-151.

Bolt, N. 2010. Pros and cons of remote usability testing. Johnny Holland Magazine. URL: http://johnnyholland.org/2010/06/pros-and-cons-of-remote-usability-testing. Accessed: 29 April 2016.

Brooke, J. 1996. SUS: a "quick and dirty" usability scale. In Jordan, P., Thomas, B., Weerdmeester, B. & McClelland, I. Usability Evaluation in Industry, pp. 189-194. Taylor & Francis. London.

Brooke, J. 2013. SUS: a retrospective. Journal of Usability studies, 8, 2, pp. 29-40. URL: http://uxpajournal.org/wp-content/uploads/pdf/JUS_Brooke_February_2013.pdf. Accessed: 15 March 2015.

Brierley, M. 2014. Usability testing: when to use remote usability testing. Spotless. Our insights. London. URL: http://www.spotless.co.uk/insights/remote-usability-testing. Accessed: 24 April 2016.

Cerejo, L. 2016. Noah’s transition to mobile usability testing. Smashing Magazine. Freiburg. URL: https://www.smashingmagazine.com/2016/02/mobile-usability-testing. Accessed: 29 April 2016.

Cocton, G., Woolrych, A. & Lavery, D. 2007. Inspection-based evaluations. In Jacko, J. & Sears, A. The Human-Computer Interaction Handbook: fundamentals, evolving technologies and emerging applications, 2d ed., pp. 1171-1188. CRC Press. Taylor & Francis Group, LLC. Boca Raton.

Davids, M., Chikte, U. & Halperin, M. 2013. An efficient approach to improve the usability of e-learning resources: the role of heuristic evaluation. In Carroll, R. Advances in Physiology Education, 37, 3, pp. 242-248. URL: http://advan.physiology.org/content/37/3/242.full.pdf+html. Accessed: 03 March 2016.

Dubey, S. & Saxena, S. 2013. Impact of software design aspects on usability. International Journal of Computer Applications, 61, 22, pp.48-53.

Dumas, J. & Fox, J. 2009. Usability testing: current practice and future directions. In Jacko, J. & Sears, A. The Human-Computer Interaction Handbook: fundamentals, evolving technologies and emerging applications, 2d ed., pp. 231-249. CRC Press. Taylor & Francis Group, LLC. Boca Raton.

Dumas, J. & Reddish, J. 1999. A Practical Guide to Usability Testing. Revised ed. Intellect Ltd. Portland.

Elling, S., Lentz, L. & Jong, M. 2011. Retrospective think-aloud method: using eye movements as an extra cue for participants' verbalizations. In CHI '11: proceedings of the SIGCHI conference on human factors in computing systems, pp. 1161-1170.

Fidas C., Katsanos, C., Papachristos, E., Tselios, N. & Avouris, N. 2007. Remote usability evaluation methods and tools: a survey. In Papatheodorou,T., Christodoulakis, D. & Karanikolas, N. Proceedings of PCI 2007, pp. 151-163. Patras. URL: http://hci.ece.upatras.gr/images/Papers/c126_remot_usability_evaluation_survey.pdf. Accessed: 11 April 2016.

Finstad, K. 2006. The system usability scale and non-native English speakers. Journal of Usability studies, 1, 4, pp. 185-188. URL: http://uxpajournal.org/wp-content/uploads/pdf/finstad_sus_non_native_speakers.pdf. Accessed: 30 February 2015.

Fluckiger, M. & Richter, M. 2014. User-centered engineering: creating products for humans. Springer-Verlag. Berlin.

Georgsson, M., Staggers, N. & Weir, C. 2016. A modified user-oriented heuristic evaluation of a mobile health system for diabetes. Self-management support. In Computers, Informatics, Nursing, 34, 2, pp. 77-84.

Grossman, T., Fitzmaurice, G. & Attar, R. 2009. A survey of software learnability: metrics, methodologies and guidelines. In CHI '09: proceedings of the SIGCHI conference on human factors in computing systems, pp. 649-658. Boston.

Guo, F. 2012. More than usability: the four elements of user experience, part I. UXmatters. URL: http://www.uxmatters.com/mt/archives/2012/04/more-than-usability-the-four-elements-of-user-experience-part-i.php. Accessed: 11 April 2016.

Haak, M., Jong, M. & Schellens, P. 2003. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. In Behaviour & Information Technology, 22, 5, pp. 339-351.

Hanington, B. & Martin, B. 2012. Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers. Beverly.

Hassenzahl, M. 2008. User experience (UX): towards an experiential perspective on product quality. Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08. Metz. ResearchGate. URL: https://www.researchgate.net/publication/238472807_User_experience_UX_Towards_an_experiential_perspective_on_product_quality. Accessed: 05 April 2016.

Hayat, H., Lock, R. & Murray, I. 2015. Measuring software usability. BCS Software Quality Management Conference, Loughborough University. Leicestershire. URL: https://dspace.lboro.ac.uk/2134/18275. Accessed: 22 March 2016.

Hodgson, P. 2010. Usability Test Data. Userfocus ltd. URL: http://www.userfocus.co.uk/articles/datathink.html. Accessed: 04 May 2016.

Hollingsed, T. & Novick, D. 2007. Usability inspection methods after 15 years of research and practice. UTEP. Departmental Papers (CS). El Paso. URL: http://digitalcommons.utep.edu/cs_papers/16. Accessed: 01 March 2016.

Holzinger, A. 2005. Usability engineering methods for software developers. In Communications of the ACM, 48, 1, pp.71-74.

Hughes, M. 2011. Reliability and Dependability in Usability Testing. UXmatters. URL: http://www.uxmatters.com/mt/archives/2011/06/reliability-and-dependability-in-usability-testing.php. Accessed: 13 May 2016.

Isbister, K. & Schaffer, N. 2008. Game usability: advancing the player experience. CRC Press. Elsevier Inc. Burlington.

ISO 9241-11:1998. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. Technical report, International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en Accessed: 25 March 2015.

ISO/IEC 25063:2014. Systems and software engineering. Systems and software product quality requirements and evaluation (SQuaRE). Common Industry Format (CIF) for usability: Context of use description. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:25063:ed-1:v1:en. Accessed: 31 May 2015.

Jeng, J. 2005. What is usability in the context of the digital library and how can it be measured? Information Technology and Libraries, 24, 2, pp. 3-12. URL: http://ejournals.bc.edu/ojs/index.php/ital/article/view/3365/2975. Accessed: 08 April 2016.

Kolski, C., Mahatody,T. & Sagar, M.2007. Cognitive walkthrough for HCI evaluation: basic concepts, evolutions and variants, research issues. EAM'07. European Annual Conference on Human-Decision Making and Manual Control. Lyngby. URL: https://www.researchgate.net/publication/228938202_Cognitive_Walkthrough_for_HCI_ev

aluation_basic_concepts_evolutions_and_variants_research_issues. Accessed: 09 March 2016.

Kortum, P. & Peres, C. 2015. Evaluation of home health care devices: remote usability assessment. JMIR Human Factors, 2, 1, e10. Toronto. URL: http://humanfactors.jmir.org/2015/1/e10. Accessed: 15 March 2016.

Krug, S. 2014. Don't make me think, revisited: a common sense approach to web usability. 3rd ed. New Riders. Berkeley.

Lewis, C., Polson, P., Rieman, J. & Wharton, C. 1994. The cognitive walkthrough method: A practitioner's guide. Institute of cognitive science. University of Colorado. Colorado. URL: http://www.colorado.edu/ics/sites/default/files/attached-files/93-07.pdf. Accessed: 11 March 2016.

Lewis, J. & Sauro, J. 2009. The Factor Structure of the System Usability Scale. Human Centered Design. Springer Berlin Heidelberg, 2009, pp. 94-103.

Macefield, R. 2014. An overview of expert heuristic evaluations. UXmatters. URL: http://www.uxmatters.com/mt/archives/2014/06/an-overview-of-expert-heuristic-evaluations.php. Accessed: 21 April 2016.

Matera, M., Rizzo, F. & Carughi, G. 2006. Web usability: principles and evaluation methods. In Mendes, E. & Mosley, N. Web Engineering, pp.143-180. Springer-Verlag. Berlin Heidelberg.

McDonald, S. & Petrie, H. 2013.The effect of global instructions on think-aloud testing. In CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2941-2944. Paris.

Moreno, A. & Yague, A. 2012. Agile user stories enriched with usability. In Wolhin, C. Agile Processes in Software Engineering and Extreme Programming: 13th International Conference, XP2012, pp.168-176. Springer-Verlag. Berlin.

Morville, P. 2004. User experience design. Semantic Studios. URL: http://semanticstudios.com/user_experience_design/. Accessed: 11 March 2015.

Nielsen, J. 1995a. Summary of usability inspection methods. Nielsen Norman Group. URL: https://www.nngroup.com/articles/summary-of-usability-inspection-methods. Accessed: 01 March 2016.

Nielsen, J. 1995b. 10 usability heuristics for user interface design. Nielsen Norman Group. URL: https://www.nngroup.com/articles/ten-usability-heuristics. Accessed: 08 March 2016.

Nielsen, J. 2000. Why you only need to test with 5 users. Nielsen Norman Group. URL: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users. Accessed: 25 February 2015.

Nielsen, J. 2001. Usability Metrics. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-metrics. Accessed 13 May 2016.

Nielsen, J. 2009. Discount Usability: 20 Years. Nielsen Norman Group. URL: https://www.nngroup.com/articles/discount-usability-20-years. Accessed: 20 March 2016.

Nielsen, J. 2010. Involving stakeholders in user testing. Nielsen Norman Group. URL: https://www.nngroup.com/articles/stakeholders-and-user-testing. Accessed 15 March 2015.

Nielsen, J. 2012a. How Many Test Users in a Usability Study? Nielsen Norman Group. URL: https://www.nngroup.com/articles/how-many-test-users. Accessed: 03 March 2015.

Nielsen, J. 2012b. Thinking aloud: The #1 Usability Tool. Nielsen Norman Group. Fremont. URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool. Accessed: 02 March 2015.

Nielsen, J. 2012c. Traveling usability lab. Nielsen Norman Group. URL: https://www.nngroup.com/articles/traveling-usability-lab. Accessed: 04 April 2016.

Nielsen, J. 2012d. Usability 101: introduction to usability. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability. Accessed 02 March 2015.

Nielsen, J. & Norman, D. The Definition of User Experience. Nielsen Norman Group. URL: https://www.nngroup.com/articles/definition-user-experience/. Accessed: 04 April 2016.

Nielsen, J. & Sova, D 2003. 234 tips and tricks for recruiting users as participants in usability studies. Nielsen Norman Group. URL: http://media.nngroup.com/media/reports/free/How_To_Recruit_Participants_for_Usability_Studies.pdf. Accessed: 25 February 2015.

NIST. 2007. Common Industry Specification for Usability – Requirements. NISTR-7432. National Institute of Standards and Technology. Information Access Division. Information Technology Laboratory. Gaithersburg. URL: http://zing.ncsl.nist.gov/iusr/documents/CISU-R-IR7432.pdf. Accessed: 22 February 2015.

Open Broadcaster Software. URL: https://obsproject.com. Accessed: 05 March 2015.

Perfetti, C. 2003. Usability Testing Best Practices: An Interview with Rolf Molich. User Interface Engineering. URL: https://articles.uie.com/molich_interview. Accessed: 03 April 2016.

Perfetti, C. 2010. Are you really prepared for your usability study? The three steps for success. User Interface Engineering. URL: https://articles.uie.com/usability_testing_three_steps. Accessed: 03 April 2016.

Perfetti, C. & Landesman, L. 2001. Eight is not enough. User Interface Engineering. URL: https://www.uie.com/articles/eight_is_not_enough. Accessed: 04 February 2015.

Petrie, H. & Bevan, N. 2009. The evaluation of accessibility, usability and user experience. In Stepanidis, C. The Universal Access Handbook. CRC Press.URL: http://www.nigelbevan.com/papers/The_evaluation_of_accessibility_usability_and_user_experience.pdf. Accessed: 07 April 2016.

Plantak, V., Kirinic, V. & Klicek, B. 2010. A Comparison of Usability Evaluation Methods for e-Learning Systems. In Katalinic, B. DAAAM International Scientific Book 2010, pp.271-288. DAAAM International. Vienna. URL: http://www.daaam.info/Downloads/Pdfs/science_books_pdfs/2010/Sc_Book_2010-027.pdf. Accessed: 02 March 2016.

Pratas, A. 2014. Creating flat design websites. E-Book. Packt Publishing. Birmingham.

Quesenbery, W. 2004. Balancing the 5Es: usability. Cutter IT Journal, 17, 2, pp. 4-11. Cutter Consortium. Arlington. URL: http://www.wqusability.com/articles/5es-citj0204.pdf. Accessed: 22 February 2015.

Rauschenberger, M., Schrepp, M., Perez-Cota, M., Olschner, S. & Thomaschewski, J. 2013. Efficient measurement of the user experience of interactive products. How to use the user experience questionnaire (UEQ). Example: spanish language version. International Journal of Interactive Multimedia and Artificial Intelligence, 2, 1, pp. 39-45. URL: http://www.measuringu.com/blog/seq10.php. Accessed: 04 March 2016.

Reis, E. 03 August 2009. Venture Hacks interview: "What is the minimum viable product?" The Lean startup blog. Startup lessons learned by Eric Ries. URL: http://www.startuplessonslearned.com/2009/03/minimum-viable-product.html. Accessed: 23 April 2016.

Robier, H. 22 March 2015. Usability versus user experience. LinkedIn post. URL: https://www.linkedin.com/pulse/usability-versus-user-experience-hannes-robier. Accessed: 11 April 2016.

Rubin J. & Chisnell, D. 2008. Handbook of usability testing: how to plan, design, and conduct effective tests. 2nd ed. Wiley Publishing, Inc. Indianapolis.

Sauro, J. 2010. If you could only ask one question, use this one. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/single-question.php. Accessed: 27 February 2015.

Sauro, J. 2011a. Are both positive and negative items necessary in questionnaires? MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/positive-negative.php. Accessed: 27 February 2015.

Sauro, J. 2011b. Measuring usability with the System Usability Scale (SUS). MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/sus.php. Accessed: 27 Feb 2015.

Sauro, J. 2011c. What’s the difference between a Heuristic Evaluation and a Cognitive Walkthrough? MeasuringU. Measuring Usability LLC. URL:http://www.measuringu.com/blog/he-cw.php. Accessed: 18 April 2016.

Sauro, J. 2012a. Comparison of usability testing methods. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/method-comparison.php. Accessed: 31 March 2016.

Sauro, J. 2012b. 8 Advantages of Standardized Usability Questionnaires. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/standardized-usability.php. Accessed: 25 February 2015.

Sauro, J. 2012c. 7 core ideas about personas and the user. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/personas-ux.php. Accessed: 25 February 2015.

Sauro, J. 2012d. 10 things to know about the single ease question (SEQ). MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/seq10.php. Accessed: 27 February 2015.

Sauro, J. 2012e. 10 things to know about unmoderated usability testing. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/unmoderated-things.php. Accessed: 31 March 2016.

Sauro, J. 2012f. 10 tips for your next survey. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/survey-tips.php. Accessed: 14 April 2016.

Sauro, J. 2013a. How to Measure Learnability. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/measure-learnability.php. Accessed: 11 May 2016.

Sauro, J. 2013b. Rating the Severity of Usability Problems. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/rating-severity.php. Accessed: 10 May 2016.

Sauro, J. 2013c. Seven tips for writing usability task scenarios. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/task-tips.php. Accessed: 09 May 2016.

Sauro, J. 2013d.10 Things to Know About the System Usability Scale (SUS). MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/10-things-SUS.php. Accessed: 09 May 2016.

Sauro, J. 2014a. 7 reasons usability problems don’t get fixed. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/problems-fixed.php. Accessed: 18 April 2016.

Sauro, J. 2014b. What to do when you can't run a usability test. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/usability-alternatives.php. Accessed: 18 April 2016.

Sauro, J. 2015a. Customer analytics for dummies. John Wiley & Sons. Inc. Hoboken.

Sauro, J. 2015b. 5 types of usability tests. MeasuringU. Measuring Usability LLC. URL:http://www.measuringu.com/blog/five-types-usability.php. Accessed: 03 April 2016.

Sauro, J. 2015c. The expert review is more than a second-rate usability test. MeasuringU. Measuring Usability LLC. URL: http://www.measuringu.com/blog/expert-review.php. Accessed: 20 April 2016.

Sauro, J. & Lewis, J. 2012. Quantifying the user experience: practical statistics for user research. Morgan Kaufmann. Waltham.

Schade, A. 2013a. Remote usability tests: moderated and unmoderated. Nielsen Norman Group. URL:https://www.nngroup.com/articles/remote-usability-tests. Accessed: 30 March 2016.

Schade, A. 2013b. Competitive usability evaluations: learning from your competition. Nielsen Norman Group. URL: https://www.nngroup.com/articles/competitive-usability-evaluations. Accessed: 27 January 2016.

Shneiderman, B. 2010. The eight golden rules of interface design. University of Maryland. University of Maryland, College Park. URL: http://www.cs.umd.edu/~ben/goldenrules.html. Accessed: 01 March 2016.

Silver, M. 2005. Exploring interface design.1st ed. Thomson/Delmar Learning. New York.

Six, J. & Macefield, R. 2016. How to determine the right number of participants for usability studies. UXmatters. URL: http://www.uxmatters.com/mt/archives/2016/01/how-to-determine-the-right-number-of-participants-for-usability-studies.php#sthash.UT5O4oXZ.dpuf. Accessed: 04 May 2016.

Soucy, K. 2010. Unmoderated, remote usability testing: good or evil? UXmatters. URL: http://www.uxmatters.com/mt/archives/2010/01/unmoderated-remote-usability-testing-good-or-evil.php. Accessed: 31 March 2016.

Spool, J. 2005. Seven common usability testing mistakes. User Interface Engineering. URL:https://www.uie.com/articles/usability_testing_mistakes. Accessed: 04 March 2015

Spool, J. 2007a. The difference between usability and user experience. User interface engineering. URL: https://www.uie.com/brainsparks/2007/03/16/the-difference-between-usability-and-user-experience. Accessed: 01 March 2016.

Spool, J. 2007b. Usability testing. Oh, the things you can learn. User Interface Engineering. URL:https://www.uie.com/articles/usability_tests_learn. Accessed: 04 March 2015.

Spool, J. 2009a. Moderating with multiple personalities: 3 roles for facilitating usability tests. User Interface Engineering. URL: https://articles.uie.com/moderating_multiple_personalities. Accessed: 03 April 2016.

Spool, J. 2009b.The $300 million button. User Interface Engineering. URL: http://www.uie.com/articles/three_hund_million_button. Accessed: 04 March 2015.

Spool, J. & Schroeder, W. 2001. Testing web sites: five users is nowhere near enough. In CHI EA '01: proceeding of CHI '01 Extended Abstracts on Human Factors in Computing Systems, pp. 285-286.

Svee, E. & Zdravkovic, J. 2014.Using intentional modeling to discover the user preferences in existing software systems. In Dalpiaz, F. & Horkoff, J. CEUR workshop proceedings, 1157. iStar 2014 Seventh International i* workshop. Thessaloniki. URL: http://ceur-ws.org/Vol-1157/paper19.pdf. Accessed: 11 March 2016.

Tho, C. 2015. The importance of minimum viable product (MVP) for startups. Tech in Asia. Singapore. URL: http://ceur-ws.org/Vol-1157/paper19.pdf. Accessed: 23 April 2016.

Thompson, K., Rozanski, E. & Haake, A. 2004. Here, there, anywhere: remote usability testing that works. In CITC5 '04: Proceedings of the 5th conference on Information technology education, pp. 132-137.

Travis, D. 2003. Discount Usability: Time to Push Back the Pendulum? Userfocus ltd. http://www.userfocus.co.uk/articles/discount.html. Accessed: 23 March 2015.

Travis, D. 2007. How to run an unmoderated usability test. Userfocus ltd. http://www.userfocus.co.uk/articles/unmoderated.html. Accessed: 31 March 2015.

Travis, D. 2010. Log usability tests like a pro. Userfocus ltd. URL:http://www.userfocus.co.uk/articles/datalogging.html. Accessed: 03 March 2015.

Travis, D. 2013a. My place or yours? How to decide where to run your next usability test. Userfocus ltd. URL: http://www.userfocus.co.uk/articles/common-types-of-usability-test.html. Accessed: 20 April 2016.

Travis, D. 2013b. The 1-page usability test plan. Userfocus ltd. URL: http://www.userfocus.co.uk/articles/usability_test_plan_dashboard.html. Accessed: 22 February 2015.

Travis, D. 2014. 5 provocative views on usability testing. Userfocus ltd. URL: http://www.userfocus.co.uk/articles/5-provocative-views.html. Accessed: 20 April 2016.

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C. & Bergel, M. 2002. An empirical comparison of lab and remote usability testing of Web sites. In Tullis, T. ResearchGate. URL: https://www.researchgate.net/publication/228540469_An_empirical_comparison_of_lab_and_remote_usability_testing_of_Web_sites. Accessed: 31 Mar 2016.

Walliman N., 2011. Research methods: the basics. Routledge. New York.

Wiklund, M., Kendler, J. & Strochlic, A. 2016. Usability testing of medical devices. 2d ed. CRC Press. Taylor & Fransis Group. Boca Raton.

Wilson, C. 2010. User experience re-mastered: your guide to getting the right design. Morgan Kaufmann. Burlington.

Wilson, C. 2014. User interface inspection methods. A user-centred design method. E-book. Morgan Kaufmann. Waltham.

Zazelenchuk, T. 2006. Heuristic evaluation and its alternatives. Userfocus ltd. URL: http://www.userfocus.co.uk/articles/heuristics.html. Accessed: 08 March 2016.

Zazelenchuk, T. 2008. Data collection for usability research. Userfocus ltd. SURL: http://www.userfocus.co.uk/articles/dataloggingtools.html. Accessed: 15 April 2016.

Appendices

Appendix 1. Persona CEO

Name: Joni Hämäläinen
Gender: male
Age: 45 years old
Education: MBA
Country: Finland

"I am aware of the growing level of Internet crimes and concerned about security of the company data. I would like to share securely vital company documentation with the clients. Additionally, I've noticed that my employees save important documents to Dropbox and I am worried that the sensitive data can end up in the hands of criminals as it's not protected enough".

Occupation: Accounting CEO

Company profile: Accounting and law, 7 people

Background: 9 years of management experience, worked in Ernst & Ernst as CFO and junior accountant before that.

Other info:

- Gets information about new software from IT consultant, networking.
- He reads news every day, likes to keep track on new technologies.
- Comfortable with technology, uses Internet, email, social networking, Internet-banking, accounting software on a daily basis.
- Access work documents from office and home.
- Spends 4-6 hours per day in front of the computer.
- Uses laptop (Mac) and tablet (Apple), iPhone.
- Impatient, get frustrated easily, but doesn't give up and looks for the solutions.
- Doesn't have enough time to learn new complicated software. Prefers fast implementation, values software credibility.
- Concerned about client satisfaction level. Empathize that the client's data is securely protected.

Appendix 2. Participant screening questionnaire

1. Name _____ [Continue]

2. What is your gender?

[ ] Male [Continue]

[ ] Female [Continue]

3. Which of the following ranges best describes your age?

[ ] under 20 years [Terminate]

[ ] 20–30 years [Continue]

[ ] 31–40 years [Continue]

[ ] 41–50 years [Continue]

[ ] over 50 [Continue]

4. What is your education level?

[ ] High-school diploma or less [Terminate]

[ ] Completed college/Undergraduate degree [Continue]

[ ] Bachelor's degree [Continue]

[ ] Master degree or higher [Continue]

5. What is your occupation and job title?

Occupation _____ [Continue]

Job title _____ [Continue]

[Recruiter: accept the candidates with the job title CEO/Manager]

6. What is your company size?

[ ] fewer than 10 persons [Continue]

[ ] from 11 to 50 persons [Continue]

[ ] more than 50 persons [Terminate]

7. How many hours per week do you spend on the computer?

[ ] Fewer than 20 hours per week [Terminate]

[ ] 20 hours per week or more [Continue]

8. What computer platform do you usually use?

[ ] Mac [Continue]

[ ] Windows [Continue]

[ ] Other [Continue]

[Recruiter: preferably recruit a mix]

9. What kind of security solutions do you currently use?

[ ] Free anti-virus solution [Continue]

[ ] Paid anti-virus solution [Continue]

[ ] Other free solutions [Continue]

[ ] Other paid solutions [Continue]

[ ] None [Terminate]

10. Are you willing to improve your security solutions?

[ ] Yes [Continue]

[ ] No [Terminate]

11. Have you ever used Lioncrypt software?

(If yes - specify, please, how many times)

[ ] yes

[ ] - from 1 to 5 times [Continue]

[ ] - more than 5 times [Terminate]

[ ] no [Continue]

12. Have you ever used Dropbox?

[ ] Yes [Continue]

[ ] No [Terminate]

Appendix 3. Participants testing schedule

Date	Start time	Planned end time	Participant	Observers
25.03.2015	16:00	17:00	Pilot test/ Participant 1	Observer 1, Observer 2, Observer 3
26.03.2015	10:30	11:30	Participant 2	Observer 1
27.03.2015	16:30	17:30	Participant 3	Observer 1
31.03.2015	12:00	13:00	Participant 4	Observer 2
02.04.2015	11:30	12:30	Participant 5	Observer 2

Appendix 4. Consent & recording release form

Consent & Recording Release Form

I agree to participate in the study conducted and recorded by the Ferus Bestia Oy.
I understand and consent to the use of the recording by Ferus Bestia Oy. I understand that the information and recording is for research purposes only and that my name will not be used for any other purpose.
I understand that participation in this usability study is voluntary and I agree to immediately raise any concerns or areas of discomfort during the session with the study administrator.
Please sign below to indicate that you have read and you understand the information on this form and that any questions you might have about the session have been answered.

Date: _____

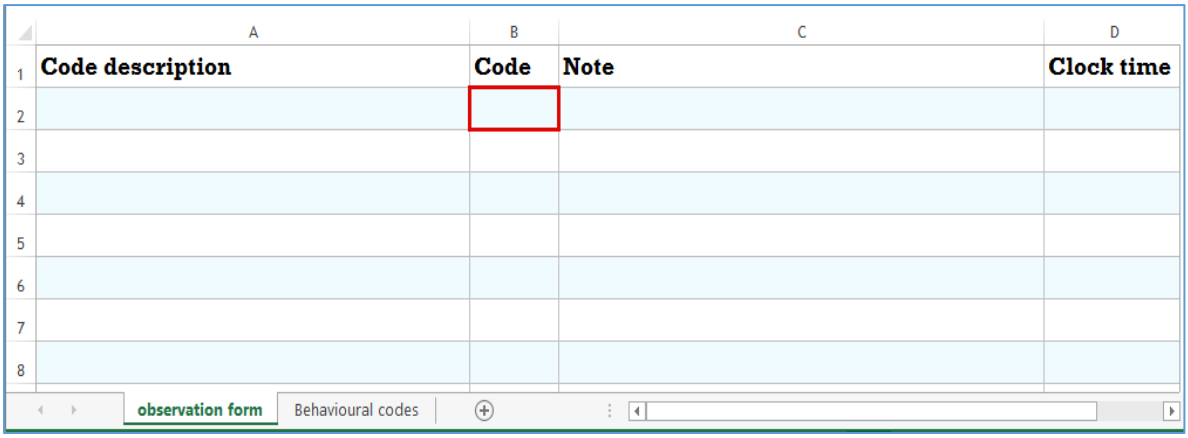
Please write your name: _____

Please sign: _____

Thank you!

We appreciate your participation.

Appendix 5. Usability test observation form



	A	B	C	D
1	**Code description**	**Code**	**Note**	**Clock time**
2				
3				
4				
5				
6				
7				
8				

observation form | Behavioural codes

Appendix 6. Logging codes

Code	Definition
A	Assist from moderator
B	Bug
C	Confused
E	End task
H	Help from moderator
M	Miscellaneous
N	Negative opinion expressed
P	Positive opinion expressed
Q	Quote, comment
S	Start task
U	User action
X	Usability problem

Appendix 7. List of tasks

N	Task	Description
1	Register	Using admin info log in on www.test.lioncrypt.com and add 2 users (save the data)
2	Log in	Log in into Lioncrypt as a new user
3	Add Dropbox	Connect to existing Dropbox account
4	Get Lioncrypt app	Download and install correct Lioncrypt app version
5	Log in to Lioncrypt app	Log into Lioncrypt app as a registered user
6	Add Lioncrypt Folder(1)	Add a new Lioncrypt Folder(1) (save the data)
7	Add file	Add 3 files to Lioncrypt Folder (1)
8	Delete file	Delete 1 file from Lioncrypt Folder (1)
9	Add Lioncrypt Folder(2)	Add a new Lioncrypt Folder(2) (save the data)
10	Add file	Add a file to Lioncrypt Folder (2) from the hard drive
11	Delete Lioncrypt Folder	Delete Lioncrypt Folder (2)
12	Share Folder	Share Lioncrypt Folder (1) with existing Lioncrypt user
extra	Save shared folder	Accept the sharing and save shared folder

Appendix 8. Participant scenario

General Scenario:

You are the CEO of a small company NOSA. Sometimes you store some crucial documents in Dropbox and you think that it might jeopardize the company's security. You would like to store encrypted files so no one can access your sensitive data.
As the company CEO you want to have control over all the company data and so you have bought a Lioncrypt “company package” **for 365 days**.

Today you need to securely share important internal documents with your company's accountant **Mr. Edward Snowden**. Use Lioncrypt software to complete the goal.

Scenarios:

Scenario I. As you remember from given instructions, **Admin** plays only an administrative role. In order to start using the Lioncrypt software you need to create **user accounts** for **yourself** and **Mr.Snowden.**

Beginning with **www.test.lioncrypt.com** and using your **Admin** account information, create **2 Lioncrypt users** (yourself and Mr. Snowden).

Admin account information:

Email: **trylioncrypt@outlook.com**

Password: **justtryit**

Your User account	**User Edward Snowden account**
Email: **ceouser@outlook.com**	Email: **mrsnowden@outlook.com**
Password _____	Password _____

Please tell us when you have finished this task. Wait for us to tell you to turn the page.

Overall, how difficult or easy did you find this task?

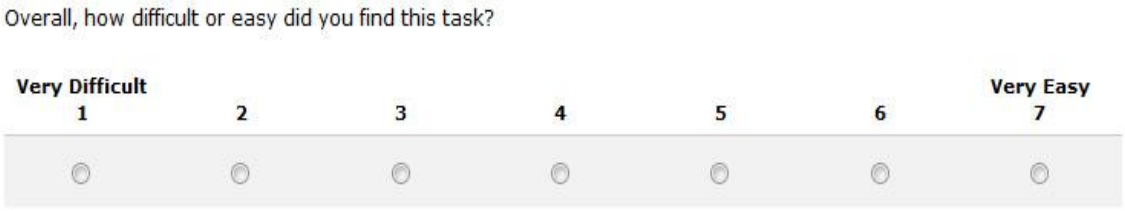
Very Difficult 1	2	3	4	5	6	Very Easy 7
○	○	○	○	○	○	○

Scenario II. Using **www.test.lioncrypt.com** connect your **User** (as the CEO) account to Dropbox.

Please tell us when you have finished this task.

Scenario III. Using **www.test.lioncrypt.com** find, download and install the appropriate version of Lioncrypt application.

Please tell us when you have finished this task. Wait for us to tell you to turn the page.



Scenario IV. Now, that you have downloaded the Lioncrypt application, create a folder for Mr. Snowden and add files you want to share.

You need to share 3 files: C://Documents/plan.txt
C://Documents/report.txt
C://Documents/private.jpg

Please tell us when you have finished this task.

Overall, how difficult or easy did you find this task?

Very Difficult 1	2	3	4	5	6	Very Easy 7
○	○	○	○	○	○	○

Scenario V. You realized that you accidentally added file **private.jpg** to the Lioncrypt Folder intended for Mr. Snowden and you don't want to share this private information with him.

Make necessary corrections.

Please tell us when you have finished this task. Wait for us to tell you to turn the page.

Overall, how difficult or easy did you find this task?

Very Difficult 1	2	3	4	5	6	Very Easy 7
○	○	○	○	○	○	○

Scenario VI. You want to store your private photos separately from Mr. Snowden's Folder, because he cannot keep secrets.

Add the following files to another folder:

C:/Documents/private.jpg

C:/Documents/private2.jpg

Please tell us when you have finished this task.

Overall, how difficult or easy did you find this task?

Very Difficult 1	2	3	4	5	6	Very Easy 7
○	○	○	○	○	○	○

Scenario VII. Your partner insisted that you delete all you private photos.

Remove your private data.

Please tell us when you have finished this task. Wait for us to tell you to turn the page.

Overall, how difficult or easy did you find this task?

Very Difficult 1	2	3	4	5	6	Very Easy 7
○	○	○	○	○	○	○

Scenario VIII. Now you are finally ready to share your top-secret documents with Snowden.

Please, complete the goal and close the application.

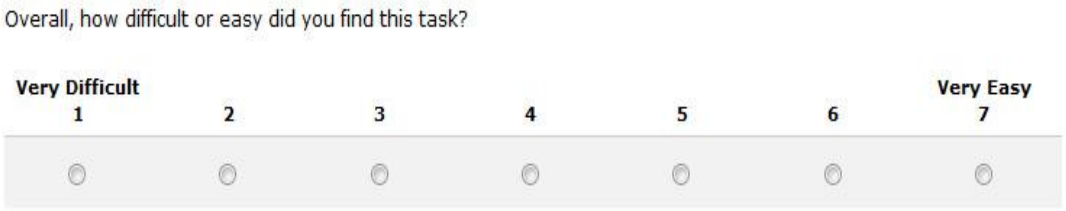
Please tell us when you have finished this task.

Please, wait until you can start this task!

Extra Scenario. Mr. Snowden has just called, he claimed you need to immediately review the documents he shared with you.

He told you the secret password: **secret**

Find out what are these important documents about.



Please tell us when you have finished this task.

All characters appearing in this scenario are fictitious. Any resemblance to real persons, living or dead, is entirely coincidental.

Appendix 9. Participants summary table

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Gender	Male	Male	Male	Male	Male
Age	20-30 years	over 50 years	20-30 years	20-30 years	41-50 years
Education level	Undergraduate degree	Master degree or higher	Undergraduate degree	Bachelor degree	Bachelor degree
Occupation /Job title	CEO	Company President and CTO	CEO	CEO	Managing director
Company size	fewer than 10 persons	fewer than 10 persons	fewer than 10 persons	fewer than 10 persons	fewer than 10 persons
Time on PC/week	20 hours per week or more	20 hours per week or more	20 hours per week or more	20 hours per week or more	20 hours per week or more
Platform	Mac	Windows	Mac	Windows	Windows
Security solutions	Free anti-virus solution	Paid anti-virus solution	Other free solutions	Free anti-virus solution	Paid anti-virus solution
Will to improve security	yes	yes	yes	yes	yes
Lioncrypt experience	yes	no	yes	no	no
How many times	from 1 to 5 times	none	from 1 to 5 times	none	none
Dropbox experience	yes	yes	yes	yes	yes

Appendix 10. Scenario success rate per each participant

	P1	P2	P3	P4	P5
Scenario 1	S	P	S	S	S
Scenario 2	S	F	S	S	S
Scenario 3	S	F	S	S	S
Scenario 4	S	F	S	P	S
Scenario 5	S	P	S	P	S
Scenario 6	S	P	S	S	S
Scenario 7	S	P	S	S	S
Scenario 8	S	S	S	S	S
Extra Scenario	S	F	S	S	P

Abbreviations:

S – success
P - partial success (1 slight hint or scenario reading)
F – failure.

P1	Participant 1
P2	Participant 2
P3	Participant 3
P4	Participant 4
P5	Participant 5

Appendix 11. Open-ended question results

Participant N	Comment
Participant 1	"When system makes you wait, could be some text about encrypting or decrypting files"
Participant 2	"It was difficult to get started. I expect easier access to sharing files and communication"
Participant 3	"Pop-ups, not at the top of the application and not bouncing all the time"
Participant 4	"User Experience and make it similar to google drive. For Dropbox connection make it green button"
Participant 5	-